Calhoun: The NPS Institutional Archive
DSpace Repository

Theses and Dissertations | 1. Thesis and Dissertation Collection, all items

1986

Parametric and nonparametric estimation of the mean number of customers in service for AN M/G/00 queue.

Park, Dong Keun

http://hdl.handle.net/10945/22020

Downloaded from NPS Archive: Calhoun

DUDLEY KNOX LIBRARY

Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

http://www.nps.edu/library

DUDLEY KNOX LIBRARY
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA 93943

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

PARAMETRIC AND NONPARAMETRIC ESTIMATION
OF THE MEAN NUMBER OF CUSTOMERS
IN SERVICE FOR AN $M/G/\infty$ QUEUE

by

Park, Dong Keun

March 1986

Thesis Advisor: P. A. Jacobs

Approved for public release; distribution is unlimited.

T226750

URITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1 REPORT SECURITY CLASSIFICATION NCLASSIFIED	1b. RESTRICTIVE MARKINGS
2 SECURITY CLASSIFICATION AUTHORITY 2 DECLASSIFICATION / DOWNGRADING SCHEDULE	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)

6 NAME OF PERFORMING ORGANIZATION aval Postgraduate School	6b OFFICE SYMBOL (*If applicable*) Code 55	7a. NAME OF MONITORING ORGANIZATION Naval Postgraduate School
ADDRESS (*City, State, and ZIP Code*) onterey, California 93943-5000		7b ADDRESS (*City, State, and ZIP Code*) Monterey, California 93943-5000
NAME OF FUNDING / SPONSORING ORGANIZATION	8b OFFICE SYMBOL (*If applicable*)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

ADDRESS (*City, State, and ZIP Code*)	10 SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO

TITLE (*Include Security Classification*)
ARAMETRIC AND NONPARAMETRIC ESTIMATION OF THE MEAN NUMBER OF CUSTOMERS IN SERVICE FOR N M/G/∞ QUEUE

PERSONAL AUTHOR(S)
ark, Dong Keun

TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (*Year, Month, Day*)	15 PAGE COUNT
aster's Thesis	FROM ______ TO ______	1986 December	73

SUPPLEMENTARY NOTATION

COSATI CODES			18 SUBJECT TERMS (*Continue on reverse if necessary and identify by block number*)
FIELD	GROUP	SUB-GROUP	M/G/∞ queueing model, simulation, Parametric estimation, Nonparametric estimation, Paired-jackknife, Separated-jackknife, Bootstrap,

ABSTRACT (*Continue on reverse if necessary and identify by block number*)

This thesis studies the estimation from interarrival and service time data of the iean number of customers in service at time t for an M/G/∞ queue. Two situations are onsidered. In one the parametric form of the service time distribution is known. In he special case in which the service time distribution is exponential the approximate ias and variance of the estimate are derived and simulation is used to study an pproximate normal confidence interval procedure. Simulation is also used to illustrate hat assuming a wrong parametric model can lead to misleading results. In the other ituation, the parametric form of the serivce time distribution is unknown and the mpirical distribution of the service times is used in the estimate of the mean number f customers in service. In the case in which the customer arrival rate is known the istribution of the estimate is derived and an approximate normal confidence interval rocedure is suggested. The use of the bootstrap and jackknife procedure to estimate

DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
a NAME OF RESPONSIBLE INDIVIDUAL Patricia A. Jacobs	22b TELEPHONE (*Include Area Code*) (408) 646-2258	22c OFFICE SYMBOL Code 55Jc

D FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

19. ABSTRACT

variability and construct confidence intervals for the estimate is also studied both analytically and by simulation.

S N 0102- LF- 014- 6601

Approved for public release; distribution is unlimited.

Parametric and Nonparametric Estimation of the Mean Number of Customers in Service for an $M/G/\infty$ Queue

by

Park, Dong Keun
Lieutenant, Republic of Korea Navy
B.S., R.O.K Naval Academy, 1979
B.S., Korea University, 1983

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 1986

Thesis D579

ABSTRACT

This thesis studies the estimation from interarrival and service time data of the mean number of customers in service at time t for an $M/G/\infty$ queue. Two situations are considered. In one the parametric form of the service time distribution is known. In the special case in which the service time distribution is exponential the approximate bias and variance of the estimate are derived and simulation is used to study an approximate normal confidence interval procedure. Simulation is also used to illustrate that assuming a wrong parametric model can lead to misleading results. In the other situation, the parametric form of the service time distribution is unknown and the empirical distribution of the service times is used in the estimate of the mean number of customers in service. In the case in which the customer arrival rate is known the distribution of the estimate is derived and an approximate normal confidence interval procedure is suggested. The use of the bootstrap and jackknife procedure to estimate variability and construct confidence intervals for the estimate is also studied both analytically and by simulation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

A. DESCRIPTION OF THE PROBLEM

The application of probability theory to a wide variety of congestion problems has been described in many papers and books [Refs. 1,2,3]. Results of queueing theory are presented in terms of component distribution functions and stochastic processes (renewal, Poisson, etc) that are taken as known; only rarely are issues addressed that arise when actual data is to be used as a basis for inference from the models; however, see Cox(1965) [Ref. 4].

The concern of this thesis is inference problems for a particularly simple queueing model, the $M/G/\infty$ queue. In this model, customers arrive according to a Poisson process with rate λ and there are an unlimited number of independent servers. Service times for each server are independent, identically distributed with distribution function F. Let $X(t)$ be the number of customers being served at time t. It is well known that if there are no customers being served at time 0, then

$$P\{X(t)=n\} = \frac{[M(t)]^k}{k!} \exp[-M(t)] \qquad (1.1)$$

where

$$M(t) = \lambda \int_0^t \bar{F}(s)ds$$

with $\bar{F}(t)=1-F(t)$ [Ref. 2]. Thus the distribution of $X(t)$ is Poisson and is characterized by its mean $M(t)$.

In this thesis we will assume that the service time distribution $F(t)$ is unknown and must be estimated from service time data and that the arrival process is known to be Poisson, except possibly for its rate . We will study the estimation of the mean number of customers being served at time t, $M(t)$.

B. SCOPE OF THE THESIS

The purpose of this thesis is to study the estimate of the mean number of customers being served at time t for a $M/G/\infty$ queue. This mean completely characterizes the distribution of the number of customers being served at time t. We will assume that the service time distribution and possibly the customer arrival rate are unknown and must be estimated from data.

We generally divide the estimation method into two cases which we shall call "parametric estimation" and "nonparametric estimation". In the parametric estimation case, a particular probabilistic model is specified for the service time distribution and the parameters of the distribution are estimated. The resulting estimate of the survivor function is then used in the estimate of the expected number of customers being served at time t. In the nonparametric estimation method, the empirical survivor function is used in the estimate of the expected number of customers.

In most cases, parametric assumptions concerning the service time distribution are difficult to justify. Hence nonparametric estimation procedure may well be preferred to parametric estimation when actual data is used. However, the nonparametric estimates can be expected to be less efficient than the parametric ones.

The thesis is organized as follows. In Chapter II, the transient distribution for the number of customers being served at time t for the $M/G/\infty$ model is described and the equilibrium distribution as time goes to infinity is obtained. In Chapter III, we study parametric estimates of the mean number of customers being served under several assumptions for service time distributions. In the special case in which the service time distribution is exponential the approximate bias and variance of the estimate are

derived and simulation is used to study an approximate normal confidence interval procedure. Parametric estimates for gamma, mixed exponential, and lognormal distributions are also considered. Simulation is used to study the effect of assuming a wrong parametric model. In Chapter IV, a nonparametric estimate of the mean number of customers being served is described. This estimate is based on the empirical distribution of the service times. In the case in which the customer arrival rate is assumed known the distribution of the nonparametric estimate is derived and an asymptotic normal confidence interval procedure is suggested. The jackknife and Bootstrap methods for obtaining approximate confidence intervals are also described. The different estimators are compared by simulation. Chapter V describes the simulation and gives the results.

In summary, this thesis studies the use of estimates of the service time distribution to obtain estimates of the mean number of customers being served for a $M/G/\infty$ queue. Both parametric and nonparametric estimates are considered and compared by simulation.

II. M/G/∞ QUEUE MODEL

The M/G/∞ queueing model is specified by the following assumptions. There are infinite number of servers. Customers arrive for service according to a Poisson process with rate λ. Service times are nonnegative independent identically distributed random variables with distribution function F. When a customer arrives, he immediately starts service.

Let X(t) represent the number of customers in service at time t. It is well known that if there are no customers being served at time 0, then

$$P\{X(t)=k\} = \frac{[\lambda p(t)]^k}{k!} \exp[-\lambda p(t)] \qquad (2.1)$$

where $p(t)=\int_0^t [1-F(s)]ds$: that is, X(t) has a Poisson distribution with mean $\lambda p(t)$ [Ref. 2]. Taking the limit as $t \to \infty$ in equation 2.1, we obtain the equilibrium distribution

$$\lim_{t\to\infty} P\{X(t)=k\} = \frac{[\lambda \int_0^\infty (1-F(x))dx]^k}{k!} \exp[-\lambda \int_0^\infty (1-F(x)dx] \qquad (2.2)$$

Thus, the limiting distribution of X(t) as $t\to\infty$ is also Poisson with mean λm, where $m=\int_0^\infty \bar{F}(x)dx$ is the mean service time. Therefore, the distribution of the number of customers being served at time t is Poisson with mean

$$M(t) = \lambda \int_0^t \bar{F}(x)dx \qquad (2.3)$$

Here, the distribution of the number of customers being served at time t is characterized by value of its mean M(t). The value of M(t) depends upon the service time distribution which is assumed unknown and must be estimated from data.

This thesis considers the problem of estimating $M(t)$ from service and interarrival time data.

III. PARAMETRIC ESTIMATION METHOD

A. DESCRIPTION

In this chapter, it will be assumed that the parametric form of the service time distribution is known. In this case the estimation of the mean number of customers being served at time t, M(t), can be considered to be a function of the parameter estimates of the distribution. In particular, the estimate of M(t), when a parametric form of the service time distribution is assumed, is denoted by $M_p(t)$, then

$$M_p(t) = \lambda \int_0^t \bar{F}(s)ds \tag{3.1}$$

where $\bar{F}(t)$ is a survivor distribution of an assumed parametric form.

In this chapter the rate of the arrival process will be assumed to be unknown. Maximum likelihood estimates of the mean interarrival times and the mean service times are used in the estimate of $M_p(t)$.

Four parametric service time distributions will be considered: the exponential, the gamma, the mixed exponential, and the lognormal distribution. In the exponential case, moment approximations are used to assess the bias of the estimate and to develop a confidence interval procedure based on asymptotic normality. The performance of the confidence interval procedure is assessed by simulation.

In the remaining three parametric models, simulation is used to assess the performance of the parametric estimates. Another source of error in using a parametric estimator is that the wrong parametric form may be used. The effect of using the (wrong) exponential model in these cases is also assessed by simulation.

Each simulation has 300 replications; each replication consists of 50 independent service times from the specified distribution, and 50 independent interarrival times from an exponential distribution. The average relative bias and the average relative square error of $M_p(t)$ are used to evaluate the performance of the parametric estimation method. All simulations were carried out on an IBM 3033 computer at the Naval Postgraduate School using the LLRANDOMII, random number generating package [Ref. 6].

B. EXPONENTIAL SERVICE TIME

In this section it will be assumed that the service time distribution is exponential; that is, $F(t) = 1-\mathrm{EXP}(-t/\mu)$, where μ is an unknown parameter and must be estimated from the observed data. The maximum likelihood estimate of μ is $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$, where x_i is the service time of the i^{th} customer. We will also assume that the rate of the Poisson arrival process λ is unknown and must be estimated.

The interarrival times of the customers are denoted by $y_1, y_2, \ldots, y_n$. Since the arrival process is Poisson with rate λ, the interarrival times are mutually independent, positive random variables with the exponential distribution function having mean $\frac{1}{\lambda}$. The maximum likelihood estimate of λ is $\hat{\lambda} = n/\sum_{i=1}^{n} y_i$. For an exponential service time distribution, an estimate of the mean number of customers in service at time t for an $M/G/\infty$ queue is

$$\hat{M}_p(t) = \hat{\lambda}\cdot\hat{\mu}\,(1-\exp[-t/\hat{\mu}]) \tag{3.2}$$

The estimate is a nonlinear function of the estimated parameters, $\hat{\lambda}$ and $\hat{\mu}$. In most cases, when estimating a function of the estimated parameters, bias is created by the nonlinear relationship of the estimated parameters. Approximate formulas for the bias and variance of $\hat{M}_p(t)$ will now be derived.

Let $\hat{\beta}$ be the mean service time, $\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} x_i$, i=1,2,..,n, and $\hat{\alpha}$ be the mean interarrival time, $\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} y_i$, i=1,2,..,n. By assumption, X_i and Y_i are independent. The estimate of $M_p(t)$ can be represented by a function of the parameters $\hat{\alpha}$ and $\hat{\beta}$ as follows:

$$M(\hat{\alpha},\hat{\beta}) = \frac{1}{\hat{\alpha}}\hat{\beta}(1-\exp[-t/\hat{\beta}]) \qquad (3.3)$$

There are no simple, exact formulas for the mean and variance of the quotient of two random variables. However, there are approximate formulas which are sometimes useful. The approximation can be obtained from the partial Taylor series expansions of $M(\hat{\alpha},\hat{\beta})$ about the true means, α and β. The expansion is

$$\begin{aligned} M(\hat{\alpha},\hat{\beta}) = {} & M(\alpha,\beta) + \frac{\partial}{\partial\alpha}M(\alpha,\beta)(\hat{\alpha}-\alpha) + \frac{\partial}{\partial\beta}M(\alpha,\beta)(\hat{\beta}-\beta) \\ & + \frac{\partial^2}{\partial\alpha\partial\beta}M(\alpha,\beta)(\hat{\alpha}-\alpha)(\hat{\beta}-\beta) + \frac{1}{2}\frac{\partial^2}{\partial\alpha^2}M(\alpha,\beta)(\hat{\alpha}-\alpha)^2 \\ & + \frac{1}{2}\frac{\partial^2}{\partial\beta^2}M(\alpha,\beta)(\hat{\beta}-\beta)^2 + R_n \qquad (3.4) \end{aligned}$$

Since we assumed that the arrival process and the service times are independent, the covariance terms turn out to be zero when we take the expectation of both sides of equation 3.4. Thus, we get

$$\begin{aligned} E[M(\hat{\alpha},\hat{\beta})] = {} & M(\alpha,\beta) + \frac{1}{2}\frac{\partial^2}{\partial\alpha^2}M(\alpha,\beta)Var(\hat{\alpha}) + \frac{1}{2}\frac{\partial^2}{\partial\beta^2}M(\alpha,\beta)Var(\hat{\beta}) \\ & + R_n \qquad (3.5) \end{aligned}$$

where R_n converges to zero at the rate $\frac{1}{n^2}$. The variance of estimate is

$$\begin{aligned} Var[M(\hat{\alpha},\hat{\beta})] = {} & [\frac{\partial}{\partial\alpha}M(\alpha,\beta)]^2 Var(\hat{\alpha}) + [\frac{\partial}{\partial\beta}M(\alpha,\beta)]^2 Var(\hat{\beta}) \\ & + R_n \qquad (3.6) \end{aligned}$$

with R_n tending to zero at the rate $\frac{1}{n^2}$. An approximate bias term, denoted by $\beta_p(t)$, can be derived immediately from the equation 3.5, that is, $\beta_p(t)=E[M(\alpha,\beta)-E[M(\hat{\alpha},\hat{\beta})]]$. Subtracting $\beta_p(t)$ from the parametric value to correct the bias, leads to the bias corrected estimate of $M_p(t)$.

In order to compare the two estimates, bias and bias-corrected, we define the following notation. Let θ_1 be the fraction of bias of $M_p(t)$ against the true value $M(t)$, and θ_2 be the fraction of square error of $M_p(t)$ against the square of the true value: that is,

$$\theta_1 = \frac{M(t) - M_p(t)}{M(t)} \tag{3.7}$$

$$\theta_2 = \left[\frac{M(t)-M_p(t)}{M(t)}\right]^2 \tag{3.8}$$

where $M(t)=\lambda\int_0^t \bar{F}(s)ds$.

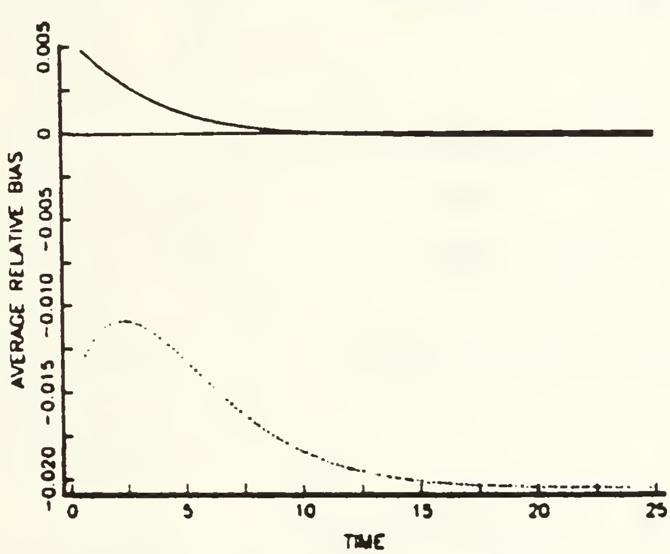

Figure 3.1 Average Relative Bias of $M_p(t)$ for Exponential with $\mu=2$ at t=1

A simulation experiment was performed to assess the performance of the estimates. In the i^{th} replication, 50 exponential interarrival times having mean 1 and 50 exponential service times having mean 2 were simulated and estimates

$$\hat{M}_p(t) = \hat{\lambda}\cdot\hat{\mu}(1-\exp[-t/\hat{\mu}]) \tag{3.9}$$

and

$$\hat{M}_p^c(t) = \hat{M}_p(t) - \beta_p(t) \tag{3.10}$$

where

$$\beta_p(t) = \frac{1}{2}\frac{\partial^2}{\partial\alpha^2}M(\alpha,\beta)\mathrm{Var}(\hat{\alpha}) + \frac{1}{2}\frac{\partial^2}{\partial\beta^2}M(\alpha,\beta)\mathrm{Var}(\hat{\beta})$$

were computed. The estimated values of $\hat{\alpha}$ and $\hat{\beta}$ were used in the variance formulas. The simulation was replicated 300 times and the average relative biases

$$\bar{\theta}_1(t) = \frac{1}{300}\sum_{i=1}^{300}\left[\frac{M(t) - \hat{M}_p^i(t)}{M(t)}\right] \tag{3.11}$$

$$\bar{\theta}_1^c(t) = \frac{1}{300}\sum_{i=1}^{300}\left[\frac{M(t) - \hat{M}_p^{ci}(t)}{M(t)}\right] \tag{3.12}$$

and the average square error

$$\bar{\theta}_2(t) = \frac{1}{300}\sum_{i=1}^{300}\left[\frac{M(t) - \hat{M}_p^i(t)}{M(t)}\right]^2 \tag{3.13}$$

$$\bar{\theta}_2^c(t) = \frac{1}{300}\sum_{i=1}^{300}\left[\frac{M(t) - \hat{M}_p^{ci}(t)}{M(t)}\right]^2 \tag{3.14}$$

were computed.

Results of the simulation are presented in Figures 3.1 and 3.2. In Figure 3.1, the dotted line shows the average

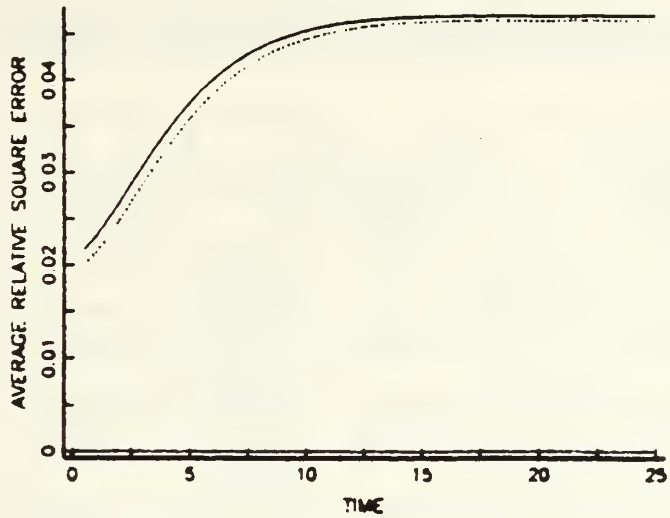

Figure 3.2 Average Relative Square Error of $M_p(t)$ for Exponential with $\mu=2$ at t=1. (n=50 , r=300)

relative bias, $\bar{\theta}_1(t)$, as a function of t for the original estimate $\hat{M}_p(t)$. The solid line shows $\bar{\theta}_1^c(t)$ for the bias corrected estimate $\hat{M}_p^c(t)$. This superimposed figure indicates that $\bar{\theta}_1^c(t)$ for the bias-corrected estimate (with solid line) is almost constant and is small. The bias estimate produces large negative value of $\bar{\theta}_1(t)$ but $\bar{\theta}_1(t)$ approaches a limiting value as $t \to \infty$. Figure 3.2 shows of the average relative square error $\bar{\theta}_2(t)$ and $\bar{\theta}_2^c(t)$ plotted as a function of time. The dotted line gives $\bar{\theta}_2(t)$ and the solid line is $\bar{\theta}_2^c(t)$. It appears from figures that the estimate of bias described in equation 3.5 does correct for the bias. However, in Figure 3.2 the bias-corrected estimate has a slightly higher relative square error than the original estimate. This higher relative square error could be due to correlation between the estimate itself and the estimate of its bias.

Simulation was used to assess the performance of the following confidence interval procedure. In the i^{th}

TABLE I

COVERAGE AND LENGTH OF 100(1-α)% C. I. FOR
THE ORIGINAL ESTIMATE (N = 50, R = 300)

trial	68 %		80 %		90 %	
	Length (s.d)	C. R.	Length (s.d)	C. R.	Length (s.d)	C. R.
1	0.2234 (0.0415)	10.67 72.00 17.33	0.2879 (0.0535)	4.00 83.00 13.00	0.3695 (0.0687)	1.00 90.00 9.00
2	0.2237 (0.0396)	9.33 73.33 17.33	0.2882 (0.0510)	4.33 81.00 14.67	0.3699 (0.0655)	1.33 86.33 12.33
3	0.2224 (0.0410)	11.00 68.67 20.33	0.2866 (0.0529)	4.33 79.00 16.67	0.3678 (0.0678)	1.33 87.33 11.33
4	0.2212 (0.0402)	9.00 66.67 24.33	0.2851 (0.0519)	3.67 76.00 20.33	0.3659 (0.0666)	1.00 83.33 15.67
5	0.2207 (0.0409)	13.33 62.33 24.33	0.2844 (0.0527)	4.67 77.67 17.67	0.3651 (0.0677)	0.67 88.67 10.67
6	0.2269 (0.0433)	11.33 70.00 18.67	0.2925 (0.0558)	3.67 82.67 13.67	0.3754 (0.0717)	0.33 89.33 10.33
7	0.2223 (0.0375)	10.67 69.00 20.33	0.2865 (0.0483)	3.00 82.00 15.00	0.3677 (0.0620)	0.33 89.67 10.00
8	0.2246 (0.0425)	7.67 71.67 20.67	0.2895 (0.0548)	2.33 81.33 16.33	0.3715 (0.0704)	0.33 87.67 12.00
9	0.2232 (0.0423)	11.00 63.67 25.33	0.2876 (0.0545)	6.00 82.67 21.33	0.3692 (0.0700)	0.67 83.67 15.67
10	0.2270 (0.0411)	13.67 67.67 18.67	0.2926 (0.0530)	3.00 81.00 16.00	0.3755 (0.0680)	0.33 87.33 12.33
Average	0.2235 (0.0410)	10.77 68.50 20.73	0.2881 (0.0528)	3.90 80.63 15.47	0.3697 (0.0678)	0.74 87.33 11.93

replication of the simulation, 50 exponential interarrival times having mean 1, and 50 exponential service times having mean 2 were generated, and $\hat{M}_p(t)$ and $\hat{M}^c_p(t)$ were computed. The approximate variance of $\hat{M}_p(t)$ was computed for t=1 using equation 3.6 with R_n=0. The 100(1-α)% confidence limits L and U were computed by

TABLE II

COVERAGE AND LENGTH OF 100(1-α)% C.I. FOR THE BIAS-CORRECTED ESTIMATE (N=50, R=300)

trial	68 %		80 %		90 %	
	Length (s.d)	C. R.	Length (s.d)	C. R.	Length (s.d)	C. R.
1	0.2234 (0.0415)	14.67 69.00 16.33	0.2879 (0.0535)	6.67 80.67 12.67	0.3695 (0.0687)	2.33 89.00 8.67
2	0.2237 (0.0396)	13.67 69.67 16.67	0.2882 (0.0510)	5.67 80.33 14.00	0.3699 (0.0655)	2.67 86.33 11.00
3	0.2224 (0.0410)	16.00 66.00 18.00	0.2866 (0.0529)	7.67 76.33 16.00	0.3678 (0.0678)	2.00 87.33 10.67
4	0.2212 (0.0402)	14.33 62.33 23.33	0.2851 (0.0519)	6.00 74.33 19.67	0.3659 (0.0666)	1.67 83.00 15.33
5	0.2207 (0.0409)	17.67 59.67 22.67	0.2844 (0.0527)	9.00 74.67 16.33	0.3651 (0.0676)	1.67 88.67 9.67
6	0.2269 (0.0433)	18.67 64.00 17.33	0.2925 (0.0558)	6.33 81.33 12.33	0.3754 (0.0717)	2.00 88.00 10.00
7	0.2223 (0.0375)	14.67 67.00 18.33	0.2865 (0.0483)	5.67 80.33 14.00	0.3677 (0.0620)	2.00 88.67 9.33
8	0.2246 (0.0425)	14.00 67.33 18.67	0.2895 (0.0548)	4.67 80.00 15.33	0.3715 (0.0704)	0.67 88.00 11.33
9	0.2232 (0.0423)	15.67 60.00 24.33	0.2876 (0.0545)	9.00 70.67 20.33	0.3692 (0.0670)	1.67 83.33 15.00
10	0.2270 (0.0411)	18.33 63.00 18.67	0.2926 (0.0530)	9.33 75.33 15.33	0.3755 (0.0680)	1.67 86.00 12.33
Average	0.2235 (0.0410)	15.77 64.80 19.43	0.2881 (0.0528)	7.00 77.40 15.60	0.3697 (0.0678)	1.84 86.83 11.33

$$L = \hat{M}_p(t) - z_{1-\frac{\alpha}{2}} \sqrt{Var[M(\hat{\alpha},\hat{\beta})]} \qquad (3.15)$$

and

$$U = \hat{M}_p(t) + z_{1-\frac{\alpha}{2}} \sqrt{Var[M(\hat{\alpha},\hat{\beta})]} \qquad (3.16)$$

where $z_{1-\frac{\alpha}{2}}$ is the upper $1-\frac{\alpha}{2}$ point of the standard normal distribution. Tables I and II show the results of 10 independent simulations for the original and the bias-corrected estimate. Each simulation was replicated 300 times. Tables report the average and standard deviation of the normal confidence interval length; the proportion, $\hat{p}$ of the intervals that covers the true value; the proportion of intervals that are too high, (e.g. $M(t) < L$); and the proportion of intervals that are too low,(e.g. $M(t) > U$).

Since the simulation replications are independent, it is possible to assess the uncertainty of $\hat{p}$. If the confidence interval procedure is correct, then p should be within approximately $\pm\ 2\sqrt{\frac{\alpha(1-\alpha)}{300}}$ of $1-\alpha$. The coverage rate in the tables indicate that the parametric estimates tend to underestimated. Obviously, the distribution of $\hat{M}_{\rho}(t)$ is skewed right. However, the confidence interval procedure works well, regardless for both the original and the bias-corrected estimate. Both have a variable coverage rate. The difference of performance between two estimates is not significant.

C. OTHER SERVICE TIME DISTRIBUTIONS

1. Mixed Exponential Service Time

In this subsection, service times having a mixed exponential parametric form will be considered. Customers arrive according to a Poisson process with unknown rate λ which must be estimated. Customers are of two types; with probability P_1 , a customer's service time is exponential with mean μ_1; with probability P_2, the customer's service time is exponential with mean μ_2. If T is the service time of an arbitrary customer, then

$$P(T > t) = P_1 \exp[-t/\mu_1] + P_2 \exp[-t/\mu_2] \tag{3.17}$$

where $P_1 + P_2 = 1$. In this case, the mean number of customers being served at time t is

$$M_p(t) = \lambda \{ P_1 \mu_1 (1 - \exp(-t/\mu_1)) + P_2 \mu_2 (1 - \exp(-t/\mu_2)) \} \quad (3.18)$$

We will assume that a customer's type and service time are observable.

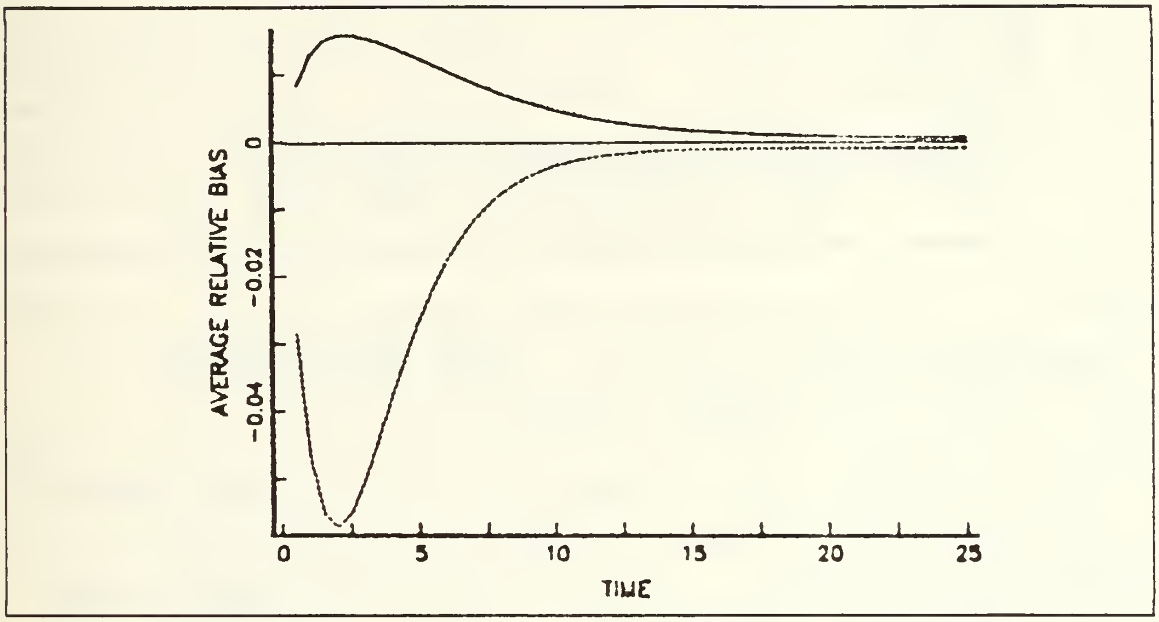

Figure 3.3 Average Relative Bias of $M_p(t)$ for Mixed Exponential with $\mu_1=2$, $\mu_2=.75$, $P_1=.2$ at t=1 (n=50, r=300)

A simulation experiment was done to assess the performance of $M_p(t)$. In the i^{th} replication, service times for 50 customers were generated, where the proportion P_1 of them have a type 1 and the proportion P_2 have a type 2. A random number was drawn to determine the type of customer. If a customer was of type i, a service time was generated from the exponential distribution with mean μ_i. For the simulation, the proportion P_1 is assumed known. The 50 independent interarrival times were also generated. In this

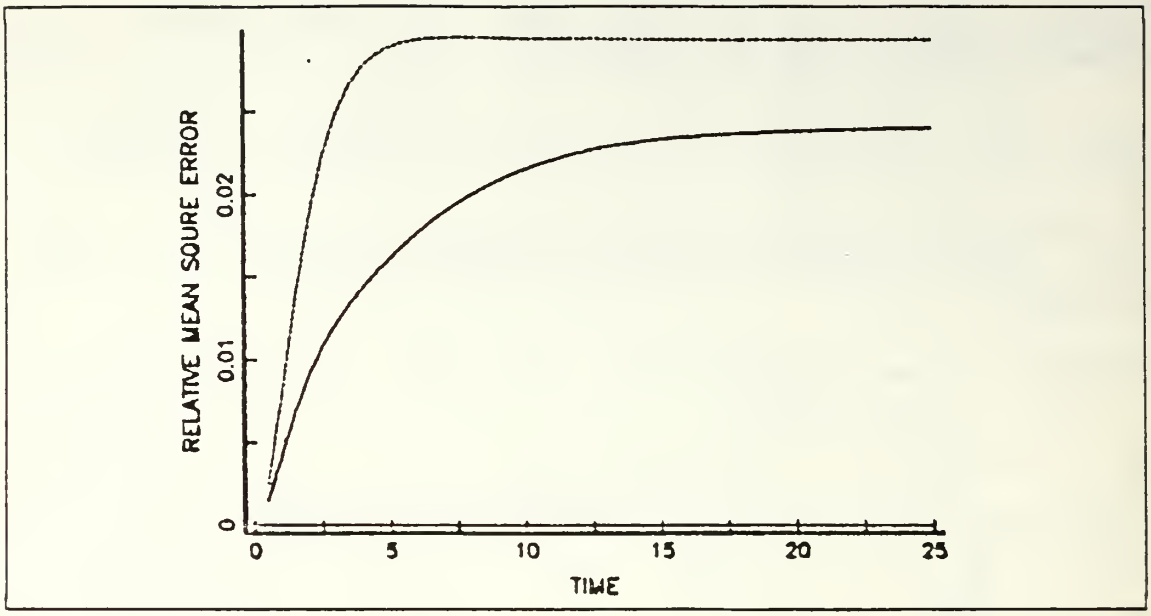

Figure 3.4 Average Relative Square Error of $M_p(t)$ for Mixed Exponential with $\mu_1=2$, $\mu_2=.75$, $P_1=.2$ at t=1 (n=50, r=300)

case , $P_1=0.2$, $\mu_1=2.$, $\mu_2=0.75$, and $\lambda=1$. The estimate $M_p(t)$ was computed; the estimate of P_i is the proportion of customers that were type i; the estimate of the mean service time μ_i was the mean service time of type i customers; the estimate of λ was the mean interarrival time. The estimate $\hat{M}_e(t)$ of $M(t)$ assuming that the service time distribution is exponential was also computed, that is,

$$\hat{M}_e(t) = \hat{\lambda}\cdot\hat{\mu}\,(1-\exp[-t/\hat{\mu}]) \qquad (3.19)$$

with $\hat{\mu}$ equal to the mean service time and $\hat{\lambda}$ is equal to the mean interarrival time. The procedure was replicated 300 times. Results of the simulation appear in Figures 3.3 and 3.4. $\bar{\theta}_1(t)$ is the average relative bias of the estimate and $\bar{\theta}_2(t)$ is the average relative square error as computed in equation 3.11 and 3.13. The solid line is for the correct parametric model estimate $\hat{M}_p(t)$. The dotted line represents the exponential model estimate $\hat{M}_e(t)$.

In both figures, we compare the correct model (represented by a solid line) with the erroneous exponential model. As expected, the exponential model clearly does not perform as well as the mixed exponential model. In Figure 3.3, the level of bias for the exponential model is very high early in time, but is reduced, and stabilizes as $t \to \infty$. Notice that the estimate using the exponential model appears too large. The average relative square error of the estimates is shown in Figure 3.4. Although the results of exponential model shows a slightly higher mean square error than that of the mixed exponential model, the results of exponential model are not too much worse. This is not surprising, since as $t \to \infty$ the estimate just depends on the mean.

2. Gamma Service Time

In this subsection, the parametric form of the service time distribution is gamma. The probability density function is

$$f(t) = \frac{1}{\Gamma(k)} \beta^k t^{k-1} e^{-\beta t} \tag{3.20}$$

where k and β are strictly positive parameters of the distribution, and k is further assumed to be an integer. By successive integrations by parts, we get

$$F(t) = 1 - \sum_{i=0}^{k-1} e^{-\beta t} \frac{(\beta t)^i}{i!} \tag{3.21}$$

its mean and standard deviation are

$$\mu = \frac{k}{\beta} \quad , \quad \sigma = \frac{\sqrt{k}}{\beta}$$

Thus, k is the parameter that specifies the degree of variability of the service times relative to the mean.

Since the arrival process is Poisson with rate λ, the interarrival times are mutually independent, positive

random variables with the distribution function $G(y)=1-EXP(-\lambda y)$, where λ must be estimated from interarrival data y_i, $i=1$ to n. Thus, the $\hat{M}_p(t)$ in this case is obtained by the successive integrations by parts of the survival function of the gamma distribution, that is

$$M_p(t) = \hat{\lambda}\left\{\frac{1}{\beta}\sum_{m=0}^{K-1}\left(1-\sum_{i=0}^{n} e^{-\beta t}\frac{(\beta t)^i}{i!}\right)\right\} \tag{3.22}$$

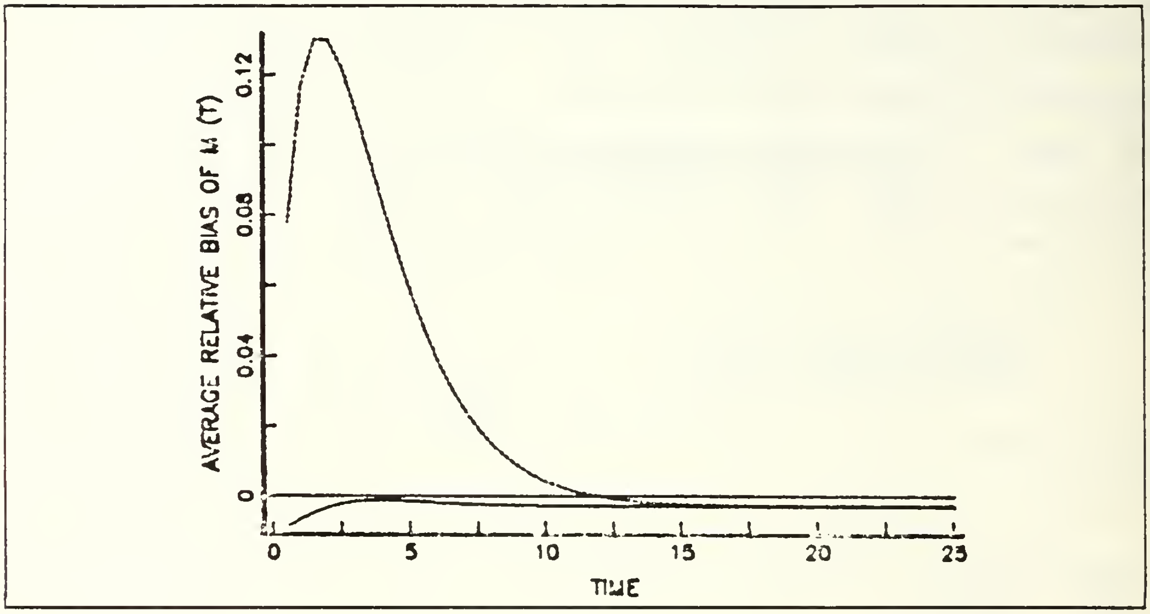

Figure 3.5 Average Relative Bias of $M_p(t)$ for Gamma with $\beta=1$, $k=2$ at $t=1$. ($n=50$, $r=300$)

The performance of $\hat{M}_p(t)$ was evaluated by the simulation. In the simulation, the parameter k of the gamma distribution was assumed to be known, but the rate of the arrival process is unknown and is estimated. Two simulation cases were run. In the i^{th} replication of the simulation, 50 independent service times were generated, where the first simulation case used the gamma distribution having $\beta=1$ and $k=2$ and the second case used the gamma distribution having $\beta=0.5$ and $k=4$; 50 independent interarrival times having $\beta=1$ were also generated. For $k=2$, the estimate is

$$\hat{M}_p(t) = \hat{\lambda}\{\frac{2}{\hat{\beta}}(1-\exp[-\hat{\beta}t] - t\cdot\exp[-\hat{\beta}t]\} \tag{3.23}$$

and for k=4, the estimate is

$$\hat{M}_p(t) = \hat{\lambda}\{\frac{4}{\hat{\beta}}(1-\exp[-\hat{\beta}t] - \exp[-\hat{\beta}t](3t + \hat{\beta}t^2 + \frac{1}{6}\hat{\beta}^2t^3)\} \tag{3.24}$$

where $\hat{\lambda} = n/\sum_{i=1}^{n} y_i$ and $\hat{\beta} = kn/\sum_{i=1}^{n} x_i$ were calculated. An estimate based on an erroneous exponential service time model parametric estimator was also calculated

$$\hat{M}_e(t) = \hat{\lambda}\cdot\hat{\mu}(1-\exp[-t/\hat{\mu}]) \tag{3.25}$$

with $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

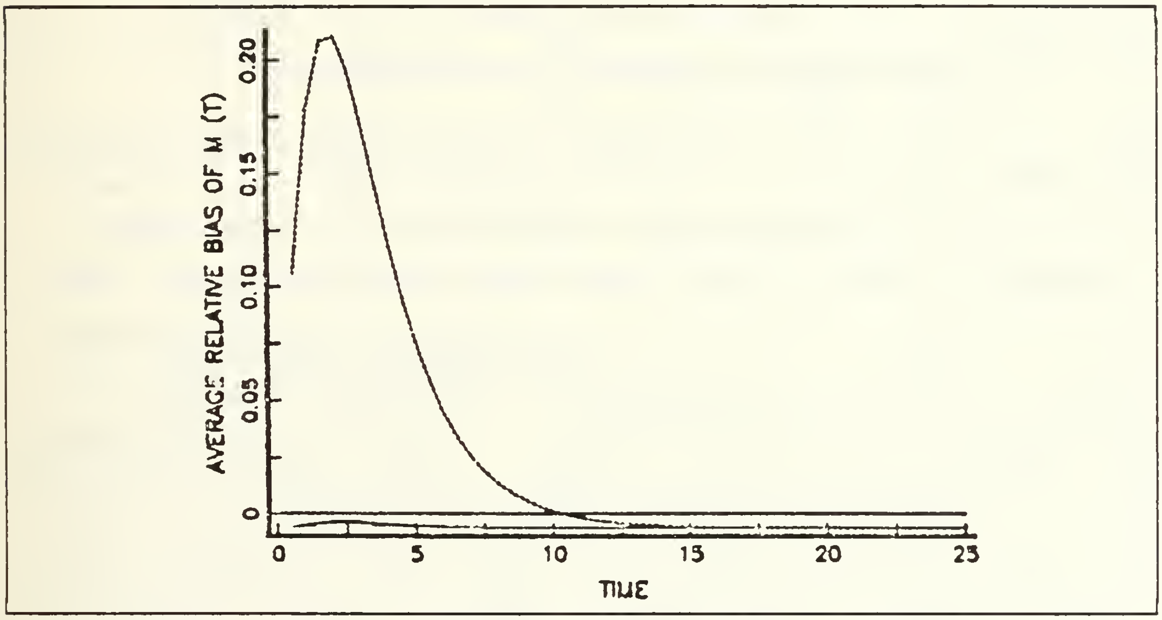

Figure 3.6 Average Relative Bias of $M_p(t)$ for Gamma with β=.5, k=4 at t=1. (n=50, r=300)

The simulation was replicated 300 times. The average relative bias $\bar{\theta}_1(t)$ and the average relative square error $\bar{\theta}_2(t)$ were calculated. These results appear in

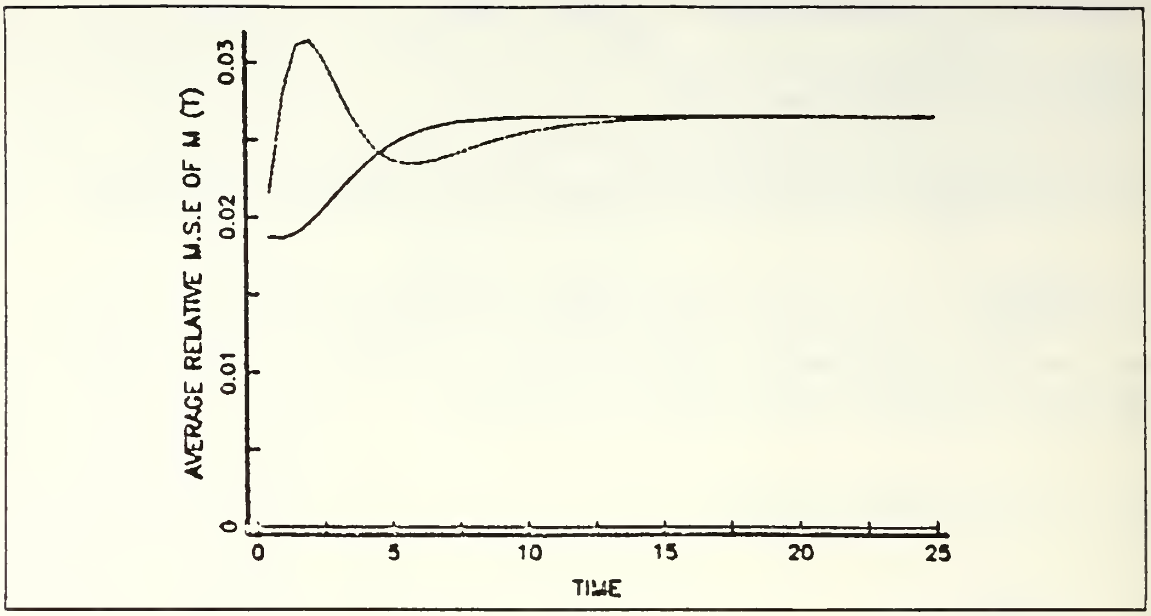

Figure 3.7 Average Relative Square Error of M (t) for Gamma with β =1, k=2 at t=1. (n=50, r=300)

Figures 3.5, 3.6, 3.7, and 3.8. The dotted lines in the figures show the results of the erroneous exponential model. The solid line represents the results of the gamma model. The figures clearly show that the performance of the exponential model is not as good as that of the gamma model initially. However, both models have the same equlibrium state for t > 15, approximately. Figure 3.5 shows the average relative bias of the estimate in the case of k=2 and Figure 3.6 shows the same in the case of k=4. In both figures, the erroneous exponential model has a high level of bias and the bias of the gamma model is almost constant; however, both models have exactly the same limiting value as $t \to \infty$. The average relative square error appears in the Figures 3.7 and 3.8. Figure 3.7 represents the average relative square error of $\hat{M}_p(t)$ and $\hat{M}_e(t)$. in case of k=2, and Figure 3.8 represents the same in the case of k=4. In Figure 3.7, the exponential model has a poorer performance than the gamma model, but the difference is small. In

Figure 3.8, the exponential model has a large value of mean square error initially and the level of mean square error associated with the exponential model grows as the value of parameter k is increased.

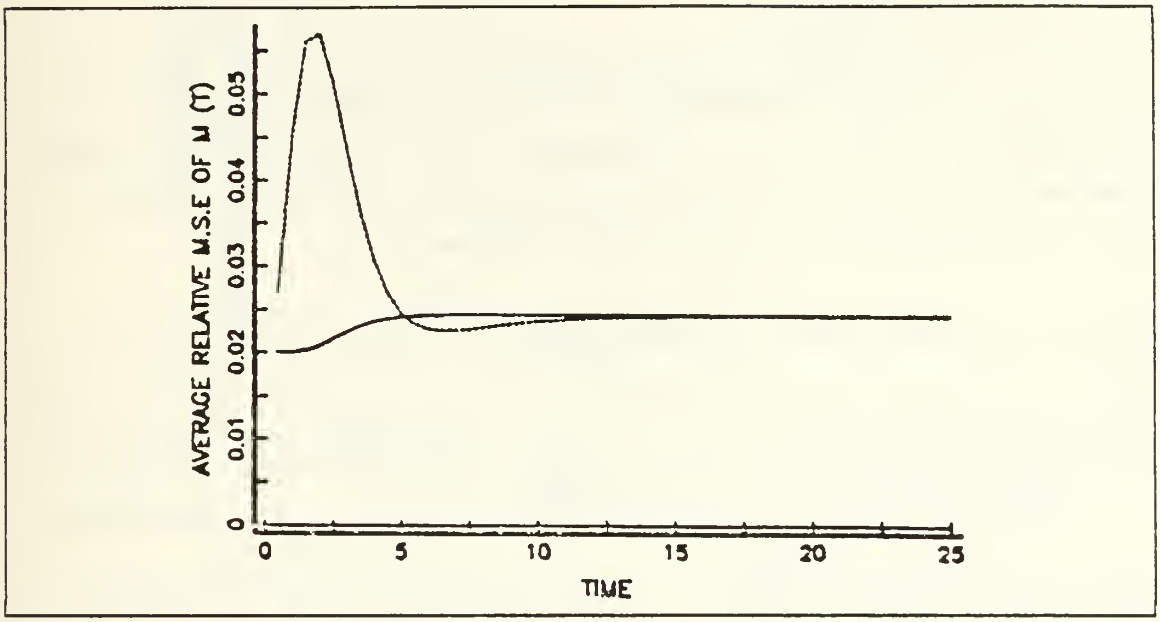

Figure 3.8 Average Relative Square Error of $M_p(t)$ for Gamma with β =.5, k=4 at t=1. (n=50, r=300)

3. Lognormal Service Time

In this subsection, the service time distribution is assumed to be lognormal. Let X be a random variable, and let a new random variable Y be defined as Y=ln X. If Y has a normal distribution, then X is said to have a lognormal distribution. The density of a lognormal distribution is given by

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2\sigma^2}(\ln x - \xi)^2\right] \qquad (3.26)$$

where $-\infty < \xi < \infty$ and $\sigma^2 > 0$. Set $Y = \ln X - \xi$ and by the

integration

$$P\{X \leq x\} = F_X(x)$$

$$= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\ln x - \xi} \exp[-y^2/2\sigma^2]\, dy$$

$$= G_Y\left(\frac{\ln x - \xi}{\sigma}\right) \qquad (3.27)$$

where $G_Y(\cdot)$ is the standard normal distribution function. If the service time has a lognormal distribution, then the estimated mean number of customers being served at time t, $\hat{M}_P(t)$, is obtained by integration-by-parts

$$\hat{M}_P(t) = \hat{\lambda}\,\{t[1-F(t)] + \int_0^t sF(s)ds\} \qquad (3.28)$$

Substituting equation 3.26 for f(t) and equation 3.27 for F(t) in the equation 3.28, we obtain

$$\hat{M}_P(t) = \hat{\lambda}\{t[1-G_Y(\frac{\ln t-\xi}{\sigma})] + \exp[\xi+\frac{\sigma^2}{2}]\,G_Y(\frac{\ln t-\xi-\sigma^2}{\sigma})\} \qquad (3.29)$$

where $G_Y(\cdot)$ is the standard normal distribution function. The lognormal distribution is positively skewed and the level of skewness depends upon the value of mean and variance of the distribution. If the value of mean is decreased but the variance is increased, then the shape of distribution tends to be more skewed and it approaches the shape of the exponential model.

The performance of the parametric estimate was assessed by simulation. In each replication of the simulation, 50 independent lognormal service times and 50 independent exponential interarrival times were generated. The simulation generated two sets of the service times. One set of service times is from the lognormal distrbution with $\xi=0.193$ and $\sigma^2=1$,and the other set is from the lognormal distribution with $\xi=0.568$ and $\sigma^2=0.5$. To estimate the mean and variance from the data, the logarithm of the service

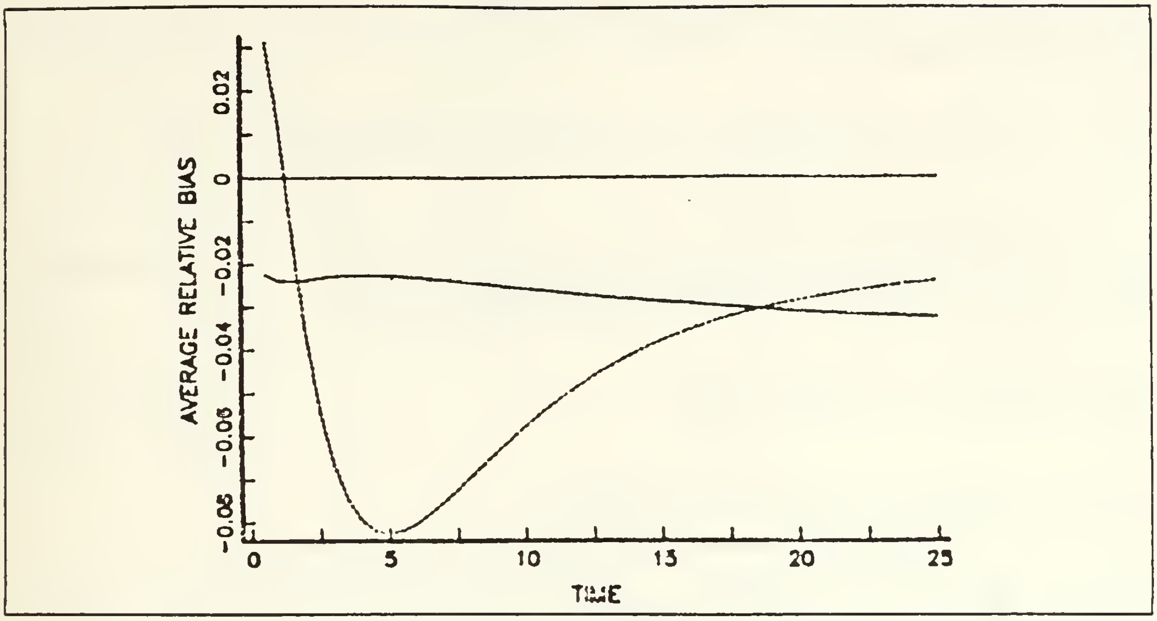

Figure 3.9 Average Relative Bias of $M_p(t)$ for Lognormal with ξ =.193, σ^2=1 at t=1. (n=50, r=300)

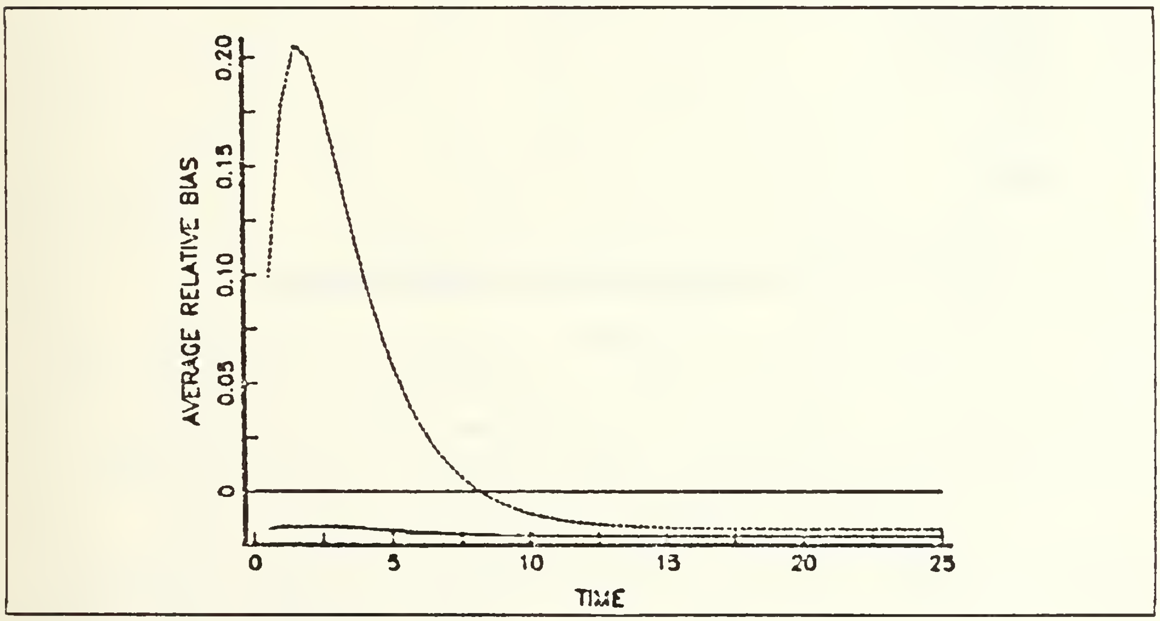

Figure 3.10 Average Relative Bias of $M_p(t)$ for Lognormal with ξ=.568, σ^2=.5 at t=1. (n=50, r=300)

time data, $t_i = \ln x_i$, i=1 to n was computed. The mean and variance are expected by

$$\hat{\xi} = \frac{1}{n}\sum_{i=1}^{n} t_i \quad , \quad \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(t_i - \hat{\xi})^2$$

Thus the estimate is

$$\hat{M}_P(t) = \hat{\lambda}\left\{ t\cdot\left[1 - G_Y\left(\frac{\ln t - \hat{\xi}}{\hat{\sigma}}\right)\right] + \exp\left(\hat{\xi} + \frac{\hat{\sigma}^2}{2}\right) G_Y\left(\frac{\ln t - \hat{\xi} - \hat{\sigma}^2}{\hat{\sigma}}\right)\right\} \qquad (3.30)$$

where $\hat{\lambda}$ is the estimate of the arrival rate. An estimate based on an erroneous exponential model

$$\hat{M}_e(t) = \hat{\lambda}\cdot\hat{\mu}\,(1 - \exp[-t/\hat{\mu}]) \qquad (3.31)$$

was also computed, where $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

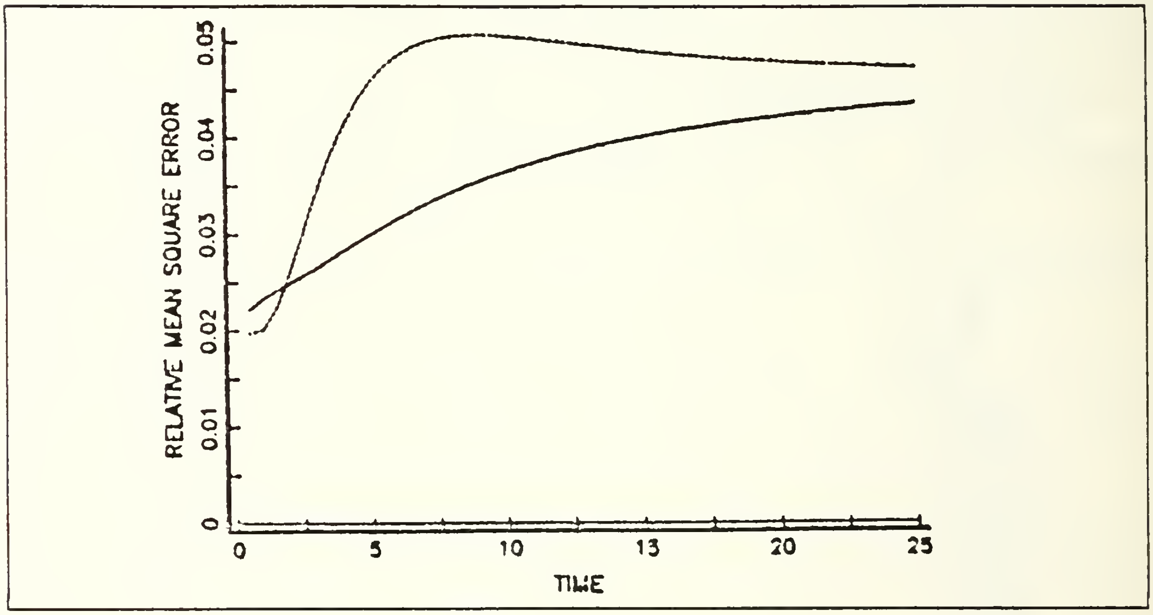

Figure 3.11 Average Relative Square Error of $M_P(t)$ for Lognormal with ξ=.193, σ^2=1 at t=1. (n=50, r=300)

The simulation was replicated 300 times. Figure 3.8 shows the tendency of $\bar{\theta}_1(t)$, the average relative bias of $\hat{M}_P(t)$, for the correct model (shown with a solid line) and the erroneous exponential model (shown with a dotted line) for the simulation with ξ=0.193 and σ^2=1 in Figure 3.9, and with ξ=0.568 and σ^2=0.5 in Figure 3.10.

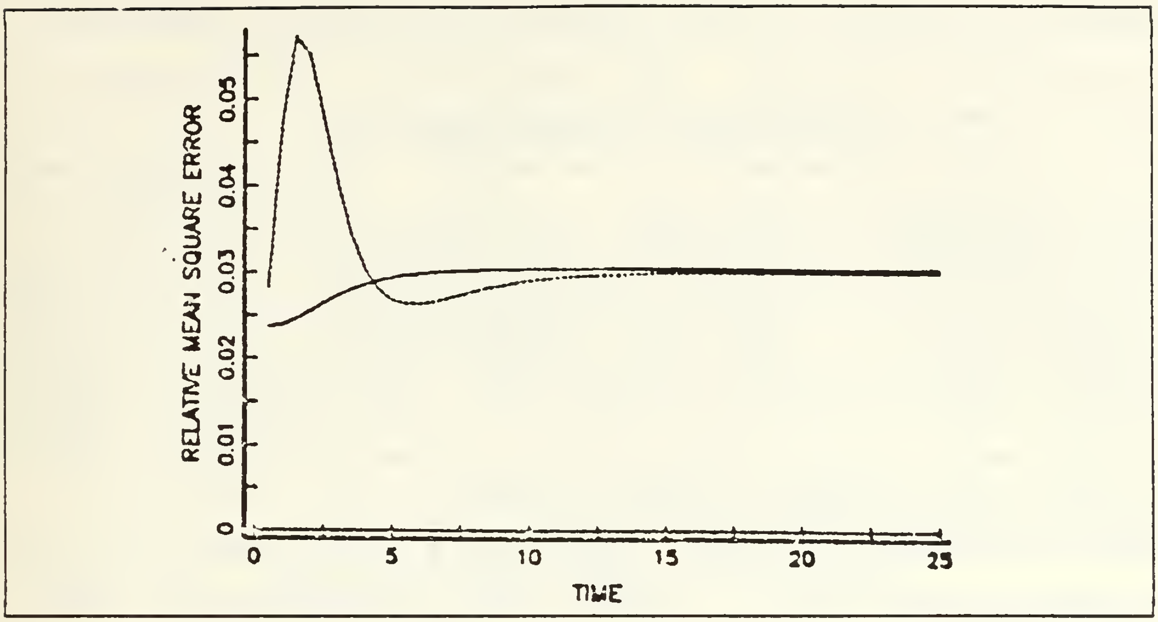

Figure 3.12 Average Relative Square Error of $M_p(t)$ for Lognormal with $\xi=.568$, $\sigma^2=.5$ at t=1. (n=50, r=300)

In both figures, the average relative bias of the exponential model is also large initially. As expected, the average relative bias of the exponential model in Figure 3.10 is larger than the results in Figure 3.9. As $t \to \infty$, the exponential model shows better performance and has a limiting value. Figures 3.11 and 3.12 show the tendency of $\bar{\theta}_2(t)$, the average relative square error, for the correct model (shown with a solid) and the erroneous exponential model (shown with a dotted line). For Figure 3.11 the parameters, $\xi=0.193$ and $\sigma^2=1$, are used to generate data. And for Figure 3.12 the parameters, $\xi=0.568$ and $\sigma^2=0.5$, are used. The value of average relative square error of the exponential model in Figure 3.12 is also higher than the results in Figure 3.11.

D. SUMMARY

The general conclusions of this chapter are that the parametric estimation method is a highly efficient for

obtaining estimates of M(t) whenever the correct assumption for the model is given. The structure of the estimate of M(t) is clearly biased since it is a nonlinear function of the estimated parameters for the service time distribution and the customer arrival rate. However, for the service time distributions considered the indications are that the bias is small. Hence the parametric estimation method performs very well, whether or not the estimate is corrected for bias, when the correct parametric form is used. However the performance of the parametric estimation is very poor when the wrong parametric model is used. For instance, the erroneous exponential model often has a high level of bias and mean-squared error. Notice that the exponential model converges to the same limiting value as the correct model as $t \to \infty$ in all the cases considered. This is because as $t \to \infty$ all estimates use the mean service time to estimate the integral of the servivor functions.

IV. NONPARAMETRIC ESTIMATION

A. DESCRIPTION

Nonparametric methods are statistical techniques which are applicable regardless of the form of the distribution function that the measurement comes from. In this chapter, these techniques will, for the most part, be based on the order statistics.

Let $x_1, x_2, \ldots, x_n$ denote a random sample from a CDF F, and let $s_{(1)}, s_{(2)}, \ldots, s_{(n)}$ denote that corresponding order statistics. Then the sample CDF is defined by

$$\hat{F}_n(t) = \frac{1}{n}(\text{number of } s_{(i)} \text{ less than or equal to t})$$

$$= \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, t]} S_{(i)}$$

For fixed time t, $\hat{F}_n(t)$ is a statistic since it is a function of the sample. In fact, for fixed time t, $\hat{F}_n(t)$ has the same distribution as that of the sample mean of a Bernoulli random variable. We know by the central limit theorem that $\hat{F}_n(t)$ is a asymptotically normally distributed with mean F(t) and variance $(\frac{1}{n})F(t)[1-F(t)]$.

Recall that (in chapter II) the mean number of customers being served at time t, M(t), is a function of the arrival rate λ and the survivor function of the service times. Hence a nonparametric estimator, denoted by $M_N(t)$, can be represented by the estimated values of $\hat{\lambda}$ and $\hat{F}(t)$. The estimated survivor function of service time is

$$\hat{\bar{F}}_n(t) = \begin{cases} 1 & \text{if } 0 \leq t < s_{(1)} \\ \frac{n-i}{n} & \text{if } s_{(i)} \leq t < s_{(i+1)} \quad \text{for } i=1,2,..,n-1 \\ 0 & \text{if } t \geq s_{(n)} \end{cases}$$

Now, using the fact that $M_N(t) = \lambda \int_0^t \bar{F}(s)ds$, we obtain a nonparametric esimate

$$\hat{M}_N(t) = \begin{cases} \hat{\lambda} t & \text{if } 0 \leq t < s \\ \hat{\lambda} [\frac{1}{n} \sum_{j=1}^{i} s_{(j)} + \frac{n-i}{n} t] & \text{if } s \leq t < s \\ \hat{\lambda} [\frac{1}{n} \sum_{j=1}^{n} s_{(j)}] & \text{if } t \geq s \end{cases} \quad (4.1)$$

where $\hat{\lambda}$ is the estimated arrival rate. Note that the nonparmetric estimate has a limiting value as $t \to \infty$, that is, $\lim_{t \to \infty} \hat{M}_N(t) = \hat{\lambda}\, m$ where m is the mean service time.

In this chapter, we will consider two different situations. In one case, we will assume that the arrival rate is known but the distribution of service times is unknown and must be estimated. Based on this assumption, $\hat{M}_N(t)$ is expressed simply in terms of the order statistics of the service times as follows

$$\hat{M}_N(t) = \lambda [\frac{1}{n} \sum_{i=1}^{\hat{K}} s + \frac{n-\hat{K}}{n} t] \quad (4.2)$$

when $s_{(K)} \leq t \leq s_{(K+1)}$. In the Appendix A, we derive the distribution of $M_N(t)$ in this case. Its mean and variance are

$$E[\hat{M}_N(t)] = \lambda \int_0^t \bar{F}(s)ds \quad (4.3)$$

$$\mathrm{Var}[\hat{M}_N(t)] = \frac{1}{n} \{ \int_0^t sF(ds) - [\int_0^t sF(ds)]^2 + t^2 F(t)\bar{F}(t)$$

$$- 2t\bar{F}(t)[\int_0^t sF(ds)] \} \quad (4.4)$$

Thus, $\hat{M}_N(t)$ is an unbiased estimate of $M_N(t)$. Further, as the sample size n is increased, $\hat{M}_N(t)$ is asymptotically normal. Thus an approximate normal $100(1-\alpha)\%$ confidence interval for $\hat{M}_N(t)$ is given by

$$\hat{M}_N(t) \pm z_{1-\frac{\alpha}{2}} \sqrt{\hat{Var}[M(t)]} \tag{4.5}$$

where $z_{1-\frac{\alpha}{2}}$ is the upper $1-\frac{\alpha}{2}$ point of standard normal distribution and $\hat{Var}[M_N(t)]$ is given in Appendix A. In this chapter, we will also study the jackknife and the bootstrap procedures for obtaining confidence intervals for $\hat{M}_N(t)$ in the case in which λ is known. In the second case, we will assume that the arrival rate λ is also unknown and must be estimated. Then, a nonparametric estimate $\hat{M}_N(t)$ is the product of two estimates,

$$\hat{M}_N(t) = \hat{\lambda}\left[\frac{1}{n}\sum_{i=1}^{\hat{K}} s + \frac{n-\hat{K}}{n} t\right] \tag{4.6}$$

where $\hat{\lambda} = n/\sum_{i=1}^{n} y_i$ and $\hat{K}$ is the number of service times that are less than or equal to t. There are no exact functional forms for the mean and variance of $\hat{M}_N(t)$ in this case. However, the jackknife and the bootstrap methods can be used to obtain confidence intervals. This will be described below.

B. JACKKNIFE ESTIMATION METHOD

In this section, we will study the jackknife procedure for obtaining a confidence interval for $M_N(t)$. The jackknife was first introduced by Quenouille (1949) for the purpose of reducing the estimate bias, and the procedure was later utilized by Tukey (1958), to develop a general method for obtaining approximate confidence intervals [Refs. 7,8].

The basic idea of the jackknife estimation method is to assess the effect of each of the groups into which the data have been divided, not by the results for that group alone, but rather through the effect upon the body of data that results from omitting that group. The two bases of the jackknife are that we make the desired calculation for all the data, and then, after dividing the data into groups, we make the calculations for each of the slightly reduced

bodies of data obtained by leaving out just one of the groups. A special case of Jackknife estimation is called the "complete jackknife estimation", where the number of subgroups is n (the size of sample); the i^{th} subgroup is obtained by deleting the i^{th} observation; thus the size of each subgroup is n-1 [Ref. 9]. Attention will be restricted to complete jackknife estimation in this study.

Let $\hat{M}^{i}_{n-1}(t)$ be the estimated mean number of customers being served at time t on the portion of the sample that omits the i^{th} sample. Let $\hat{M}_{all}(t)$ be the corresponding estimator for the entire sample and define the i^{th} Pseudo-value by

$$\hat{M}_{i}(t) = n\hat{M}_{all}(t) - (n-1)\hat{M}^{i}_{n-1}(t) \tag{4.7}$$

The jackknife estimate $\hat{M}_J(t)$ and an estimate $\hat{S}^2_J$ of its variance are given by

$$\hat{M}_J(t) = \frac{1}{n}\sum_{i=1}^{n}\hat{M}_i(t) \tag{4.8}$$

$$\hat{S}^2_J = \frac{1}{n(n-1)}\sum_{i=1}^{n}[\hat{M}_i(t) - \hat{M}_J(t)]^2 \tag{4.9}$$

Tukey (1958) proposes that the n estimated pseudo values be treated as approximately independent and identically distributed random variables [Ref. 9]. Hence, the statistic

$$\frac{\sqrt{n}\,(\hat{M}_J(t) - \hat{M}_{all}(t))}{\left[\frac{1}{n-1}\sum_{i=1}^{n}(\hat{M}_i(t) - \hat{M}_J(t))^2\right]^{1/2}} \tag{4.10}$$

has an approximate t-distribution with n-1 degrees of freedom, which leads to the approximate $100(1-\alpha)\%$ confidence interval

$$\hat{M}_J(t) \pm t_{1-\frac{\alpha}{2}}\,\hat{S}_J \tag{4.11}$$

where $t_{1-\frac{\alpha}{2}}$ is the upper $1-\frac{\alpha}{2}$ critical point of the t-distribution with n-1 degrees of freedom. The confidence interval given by equation 4.11 is a function of the estimated variance. In the remainder of this section, we will describe several methods of implementing the confidence interval procedure. We will also obtain an analytic expression for the jackknife estimate and its variance estimate for the case in which the arrival rate λ is known.

1. <u>Jackknife Estimate with Known Arrival Rate</u>

In this subsection, the arrival rate is asssumed known. In this case the closed form expression for the jackknife estimate and its variance estimate can be derived.

The nonparametric estimate of the mean number of customers being served at time t, $\hat{M}_N(t)$, can be expressed in terms of the service times as follows:

$$\hat{M}_N(t) = \lambda\left[\frac{1}{n}\sum_{i=1}^{\hat{K}} s_{(i)} + \frac{n-\hat{K}}{n}t\right] \tag{4.12}$$

where $S_{(i)}$'s are the order statistics of independent and identically distributed random quantities from the unknown probability distribution F, and the variable $\hat{K}$ is the number of $S_{(i)}$'s which are less than t. This equation shows immediately that $\hat{M}_N(t)$ is the linear function of the order statistics of service times.

The jackknife estimate is based on sequentially deleting point $S_{(i)}$ and recomputing the estimator. Removing point $S_{(i)}$ from data set gives a different empirical probability distribution $\hat{F}_{n-1}$ with mass $\frac{1}{n-1}$ at $S_{(1)}, S_{(2)}, \ldots, S_{(i-1)}, S_{(i+1)}, \ldots, S_{(n)}$ and a corresponding recomputed value of the estimate. In the jackknife process, the i^{th} pseudo value is

$$\hat{M}_i(t) = \begin{cases} \lambda t & \text{if } i > \hat{K} \\ \lambda s_{(i)} & \text{if } i \leq \hat{K} \end{cases} \tag{4.13}$$

for the fixed time t. Accordingly the pseudo values $\hat{M}_i(t)$ have just $\hat{K}+1$ different values. The jackknife estimate is

$$\hat{M}_J(t) = \lambda\left[\frac{1}{n}\sum_{i=1}^{\hat{K}} s_{(i)} - \frac{n-\hat{K}}{n}t\right] \tag{4.14}$$

This result is exactly the same as the original estimate. This is because the estimate $\hat{M}_J(t)$ is unbiased. In Appendix B, the jackknife variance estimate is derived as follows:

$$\hat{\mathrm{Var}}[M_J(t)] = \frac{1}{n-1}\left\{\frac{1}{n}\sum_{i=1}^{\hat{K}} s_{(i)}^2 - 2t\hat{\bar{F}}_n(t)\left[\frac{1}{n}\sum_{i=1}^{\hat{K}} s_{(i)}\right] - \left[\frac{1}{n}\sum_{i=1}^{\hat{K}} s_{(i)}\right]^2 + t^2\hat{F}_n(t)\hat{\bar{F}}_n(t)\right\} \tag{4.15}$$

where $\hat{\bar{F}}_n(t)$ is the sample survivor function. Comparing equation 4.15 with equation 4.3 we see that $\left(\frac{n}{n-1}\right)[\mathrm{Var}\{\hat{M}_N(t)\}] = E[\hat{\mathrm{Var}}\{M_J(t)\}]$. Thus the jackknife variance estimate tends to be conservative in the sense that its expectation is greater than the true variance of $M_N(t)$. We will now describe two selected procedures to obtain confidence intervals for the jackknife estimate. Tukey suggested that the statistic in equation 4.10 has an approximate t-distribution with n-1 degrees of freedom, which leads to the approximate two-sided 100(1-α)% confidence interval

$$\hat{M}_J(t) \pm t_{1-\frac{\alpha}{2}}\sqrt{\hat{\mathrm{Var}}[M_J(t)]} \tag{4.16}$$

for $M_N(t)$, where $t_{1-\frac{\alpha}{2}}$ is the upper $1-\frac{\alpha}{2}$ critical point of the t-distribution with n-1 degrees of freedom. However, the n estimate pseudo values have just $\hat{K}+1$ different values. Hence, another possible procedure is to adjust the degrees of freedom of the t-distribution, that is, subtract one from the number of different pseudo values ($\hat{K}+1$), and use the result as the degrees of freedom. The length of confidence interval generated using the adjusted degrees of freedom ($\hat{K}$) is slightly wider than that generated using the usual

degrees of freedom (n-1) and the coverage rate should be increased.

2. Jackknife Estimate with Unknown Arrival Rate

In this subsection, it will be assumed that the rate of Poisson arrival process is unknown and must also be estimated. The maximum likelihood estimate is $\hat{\lambda} = n/\sum_{i=1}^{n} y_i$, where y_i is the interarrival time between i^{th} and $(i-1)^{th}$ customers. A nonparametric estimate of mean number of customers being served at time t is given by

$$\hat{M}_N(t) = \hat{\lambda}\{\frac{1}{m}\sum_{i=1}^{\hat{K}} s_{(i)} + \frac{m-\hat{K}}{m}t\} \qquad (4.17)$$

where $S_{(i)}$'s are the order statistics of independent, identically distributed random quantities from the unknown probability distribution F. It is assumed that the $S_{(i)}$'s and Y_i's are independent. The variable $\hat{K}$ is the number of S 's which are less than t. The data consist of two independent random samples,

$S_1, S_2, \ldots, S_m \sim F$ and $Y_1, Y_2, \ldots, Y_n \sim Q$

F and Q being two possibly different distribution on the real line with Q, the exponential distribution with mean $\frac{1}{\lambda}$. From equation 4.17, the estimate $\hat{M}_N(t)$ is the product of two estimates. One is the function of y_i , $\hat{\lambda} = n/\sum_{i=1}^{n} y_i$, and the other is the function of s_i , $H(s) = \frac{1}{m}\sum_{i=1}^{\hat{K}} s_i + \frac{m-\hat{K}}{m}t$. there are many possible ways to perform a two-sample jackknife procedure. We will call one method the "paired sample jackknife" procedure. Since the size of both samples is the same, we make the one set of observations by pairing respective observations, that is, $(s_1, y_1), (s_2, y_2), \ldots, (s_m, y_m)$). As with the one-sample jackknife, we estimate the $\hat{M}_{all}(t)$ for all the data, and then, we estimate $\hat{M}^{\hat{i}}_{m-1}(t)$ based on the remaining data obtained by leaving out just the i^{th} pair. Thus the i^{th} pseudo value $\hat{M}_i(t)$ is

$$\hat{M}_i(t) = n\hat{M}_{all}(t) - (n-1)\hat{M}^{\hat{i}}_{m-1}(t) \qquad (4.18)$$

and the jackknife estimate $\hat{M}_J(t)$ and variance estimates are given by

$$\hat{M}_J(t) = \frac{1}{n}\sum_{i=1}^{n}\hat{M}_i(t)$$

$$\hat{S}_J^2 = \frac{1}{n(n-1)}\sum_{i=1}^{n}[\hat{M}_i(t) - \hat{M}_J(t)]^2$$

Based on these statistics, an approximate two-sided 100 $(1-\alpha)\%$ confidence interval is given by

$$M(t) \pm t_{1-\frac{\alpha}{2}}\hat{S}_J \tag{4.19}$$

where $t_{1-\frac{\alpha}{2}}$ is the upper $1-\frac{\alpha}{2}$ point of t-distribution with n-1 degrees of freedom. A second method is called the "separated sample jackknife" procedure. Since we assumed that the X_i's and Y_i's are independent, we can perform the jackknife procedure separately for each sample, and then, estimates which combine jackknife estimates and the jackknife variance estimate can be computed.

Let $\hat{M}_{J1}(t)$ be the jackknife estimate of λ and $\hat{V}_y$ be the jackknife variance estimate for λ. Let $\hat{M}_{J2}(t)$ be the jackknife estimate of $\int_0^t \bar{F}(s)ds$ and $\hat{V}_S$ be its jackknife variance estimate. Then the combined jackknife estimate of $\hat{M}_{JC}(t)$ is

$$\hat{M}_{JC}(t) = \hat{M}_{J1}(t)\cdot\hat{M}_{J2}(t) \tag{4.20}$$

and the combined jackknife variance estimate is

$$\hat{S}_{JC}^2 = \hat{V}_y\cdot\hat{V}_S + \hat{V}_y[\hat{M}_{J2}(t)]^2 + \hat{V}_S[\hat{M}_{J1}(t)]^2 \tag{4.21}$$

The approximate two-sided 100 $(1-\alpha)$ % confidence interval is given by

$$\hat{M}_{JC}(t) \pm t_{1-\frac{\alpha}{2}}\hat{S}_{JC}/\sqrt{n} \tag{4.22}$$

where $t_{1-\frac{\alpha}{2}}$ is the upper $1-\frac{\alpha}{2}$ point of t-distribution with n-1 degrees of freedom.

C. BOOTSTRAP ESTIMATION METHOD

Efron(1979) introduced the bootstrap method for estimating the distribution of a statistic computed from observations [Ref. 10]. The bootstrap estimate is obtained by replacing the unknown distribution by the empirical distribution of the data in the definition of the statistical function. In practice, the distribution of the statistic is approximated by Monte Carlo methods.

For convenience, the arrival rate is assumed to be known and equal to 1, then the nonparametric estimate $\hat{M}_N(t)$ is just a function of service times. This is a one-sample problem. The bootstrap procedure is as follows:

1. Suppose that the data points $x_1, x_2, \ldots, x_n$ are independent observations from the unknown distribution F. Then the true estimate is

$$M_N(t) = \int_0^t \bar{F}(s)ds \tag{4.23}$$

2. We can estimate the distribution F by the empirical probability distribution F_n.
   F_n : mass $\frac{1}{n}$ on each observed data point x_i $i=1,2,\ldots,n$
3. The bootstrap estimate of $M_N(t)$ is

$$\hat{M}_B(t) = \int_0^t \bar{F}_n(s)ds \tag{4.24}$$

To obtain an estimate of variability for $\hat{M}_B(t)$, we procede as follows

(1) Construct F_n, the empirical distribution function, as just described.

(2) Draw a bootstrap sample $x_1^*, x_2^*, \ldots, x_m^*$ by independent random sampling from F_n.

Notice that we are not getting a permutation distribution since the values of X_i^* are selected with replacement from

the set $(x_1, x_2, \ldots, x_n)$. As a point of comparison, the ordinary jackknife can be thought of as drawing samples of size n-1 without replacement.

(3) Compute an estimate of $M_N(t)$ for each bootstrap replication, $M_N^*(t)$, that is, the value of statistic evaluated for the bootstrap sample.

$$M_N^*(t) = \frac{1}{n} \sum_{x_i^* \leq t} x_i^* + \frac{1}{n}[n - \sum_{i=1}^{n} I(x_i^* \leq t)]\, t \tag{4.25}$$

where $I(x \leq t) = \begin{cases} 1 & \text{if } x \leq t \\ 0 & \text{otherwise} \end{cases}$

(4) Do step (2) some large number "B" times obtaining independent bootstrap replications $M_N^{*1}(t)$, $M_N^{*2}(t), \ldots, M_N^{*B}(t)$.

Based on the bootstrap replications, the approximate estimate of $M_N(t)$ and its variance are obtained by

$$\hat{M}_B(t) = \frac{1}{B} \sum_{i=1}^{B} M_N^{*i}(t) \tag{4.26}$$

$$\hat{Var}[M_B(t)] = \frac{1}{B-1} \sum_{i=1}^{B} [M_N^{*i}(t) - \hat{M}_B(t)]^2 \tag{4.27}$$

A formula for the conditional variance of $\hat{M}_B(t)$ given the original sample data is derived in Appendix C. This expression is given by

$$Var[M_B(t)] = \frac{1}{n}\left\{\frac{1}{n} \sum_{i=1}^{\hat{K}} x_i^2 - \left[\frac{1}{n} \sum_{i=1}^{\hat{K}} x_i\right]^2 + t^2\left[\frac{n-\hat{K}}{n}\right]\frac{\hat{K}}{n} - 2t\left[\frac{n-\hat{K}}{n}\right]\left[\frac{1}{n} \sum_{i=1}^{\hat{K}} x_i\right]\right\} \tag{4.28}$$

Notice that the expected value of the conditional variance of the bootstrap estimate is approximately equal to the variance of the nonparametric estimate of $M_N(t)$ which is derived in Appendix A.

So far we have considered the problem, where the arrival rate is known. The bootstarp methodology also applies if the arrival rate is unknown and is estimated from

interarrival data. Suppose the data consist of a random sample $X=(X_1, X_2, \ldots, X_n)$ from unknown service time distribution F and an independent sample $Y=(Y_1, Y_2, \ldots, Y_n)$ from the exponential interarrival time distrbution G with unknown parameter λ. One bootstrap procedure to estimate the expected number of customers being served at time t is to construct F_n and G_n, the empirical probability distribution corresponding to F and G. Bootstrap samples $X_i^* \sim F_n$, i=1,2,...n, $Y_i^* \sim G_n$, j=1,2,...n, are independently drawn, an estimate of $M_N(t)$

$$M_N^*(t) = \frac{n}{\sum_{i=1}^{n} Y_i^*}\left\{\frac{1}{n}\sum_{X_i^* \le t} X_i^* + \frac{1}{n}\left[n - \sum_{i=1}^{n} I(X_i^* \le t)\right]t\right\} \qquad (4.29)$$

is calculated. As before there are a large number B of bootstrap replications. For this case, the bootstrap estimate of $M_N(t)$ and its variance are still given by equations 4.26 and 4.27. There appears to be no closed form of the analytical variance of $\hat{M}_B(t)$ in this case. Now we will describe methods to obtain approximate confidence intervals for the bootstrap estimate $\hat{M}_B(t)$.

1. <u>The Percentile Method</u>

A simple method for assigning approximate confidence intervals to the nonparametric estimate $M_N(t)$ is as follows: Let

$$\hat{C}(t) = \frac{\#\{M_N^{*i}(t) \le t\}}{B} \qquad (4.30)$$

be the cumulative distribution function of the bootstrap distribution of $M_N(t)$; B is the number of bootstrap replications. For a given $0<\alpha<0.5$, define

$$\hat{L}(\alpha) = \hat{C}^{-1}(\alpha) \quad , \quad \hat{U}(\alpha) = \hat{C}^{-1}(1-\alpha)$$

Usually denoted simply by $\hat{L}$ and $\hat{U}$. This definition runs into complications when we actually try to compute quantiles $\hat{L}$ and $\hat{U}$ from a set of bootstrap replications. To overcome

these difficulties, we order the bootstrap replications from smallest to largest, obtaining the sorted data $M_N^{*(i)}(t)$, for i=1 to B. Letting represent any fraction between 0 and 1; take $Q(\alpha)$ to be $M_N^{*(i)}(t)$ whenever Q is one of the functions $\alpha_i = \frac{i-0.5}{B}$, for i=1 to B. Thus $\hat{L}(\alpha)$ turns out to be the $(B*\alpha+0.5)^{th}$ $M_N^{*(i)}(t)$ and $\hat{U}(\alpha)$ to be the $(B*(1-\alpha)+0.5)^{th}$ $M_N^{*(i)}(t)$. The percentile method consists of taking

$$[\ \hat{L}(\alpha)\ ,\ \hat{U}(\alpha)\ ] \tag{4.31}$$

as an approximate $1-2\alpha$ confidence interval for $\hat{M}_B(t)$ since $\alpha=C(\hat{L})$, $1-\alpha=C(\hat{U})$, the percentile method interval consists of the central $1-2\alpha$ proportion of the bootstrap distribution.

2. The Bias-corrected Percentile Method

Efron(1980) suggests the following bias correction for the percentile confidence interval procedure [Ref. 11]. He argues that if $\hat{M}_B(t)$ is not the median of the bootstrap replication distribution, then a bias correction to the percentile method is called for. To be specific, define

$$z_o = \Phi^{-1}[\hat{C}(\hat{M}_B(t))] \tag{4.32}$$

where $\hat{C}(t) = \frac{\#\{M_N^*(t) \le t\}}{B}$ as in equation 4.30, and is the cumulative distribution function for a standard normal variate. The bias corrected percentile method consists of taking

$$[\hat{C}^{-1}\{\Phi(2z_o - z_\alpha)\}\ ,\ \hat{C}^{-1}\{\Phi(2z_o + z_\alpha)\}] \tag{4.33}$$

as an approximate $1-2\alpha$ central confidence interval for $\hat{M}_B(t)$. Here z_α is the upper point for a standard normal $\Phi(z_\alpha)=1-\alpha$.

Notice that if $\hat{M}_B(t)$ is the median of the bootstrap distribution then $z_0=0$ and equation 4.33 reduces to equation 4.31, the uncorrected percentile interval. However, even small differences of $\Pr\{M^*_N(t) \leq \hat{M}_B(t)\}$ from 0.5 can make equation 4.33 much different from equation 4.31.

V. SIMULATION RESULTS

The purpose of the simulation in this chapter is to assess the performance of the nonparametric estimation methods, the jackknife and the bootstrap. Since the estimate of M(t), the mean number of customers being served at time t, is a function of the customer arrival rate and the integral of the survivor function of the service time distribution, two simulations cases are done. The first simulation case was performed to estimate $M_N(t)$, the nonparametric estimate of M(t), as a function of the service times with the arrival rate assumed to be known and set equal to 1. For this case, the jackknife and bootstrap estimate of the variance were derived in the chapter IV, and compared with the numerical estimate obtained by the simulation. The second case assumed that the customer arrival rate is also unknown and must be estimated using interarrival times.

In each replication of the simulation for case 1, 50 independent service times from a specified service time distribution were generated. For the bootstrap procedure, 500 bootstrap replications were performed. The simulation was replicated 300 times. For the purposes of comparison, we considered four types of service time distributions, which were the exponential, the mixed exponential, the gamma, and the lognormal distribution. The arrival process is known to be Poisson process with known rate $\lambda=1$. The same generated service times were used for each estimation procedure in a replication. This reduces the variability of the differences in performance between the procedures. All programming was done on IBM 3033 computer at the Naval Postgraduate school using the LLRANDOMII, random number generating package [Ref. 6].

TABLE III

STATISTICAL DATA OF ESTIMATE OF M(T)
N=50 R=300 (B=500)

	True value	Parametric		Nonparametric	
		Correct	Errors	Jack	Boot
Exponential (μ = 2)	0.7869	0.7830 (0.0268)	-	0.7820 (0.0451)	0.7819 (0.0452)
Mixed expon. (μ_1=2, μ_2=.75, p_1=.2)	0.5992	0.5871 (0.0023)	0.6246 (0.0026)	0.5991 (0.0512)	0.5989 (0.0512)
Gamma (β=1, κ=2)	0.8963	0.8952 (0.0009)	0.7861 (0.0010)	0.8965 (0.0325)	0.8967 (0.0325)
Lognormal (ξ=.193, σ^2=1)	0.8094	0.8140 (0.0021)	0.7844 (0.0020)	0.8139 (0.0383)	0.8138 (0.0383)

Table III presents the results of several estimation methods when the arrival rate is given and equal to 1. The top of each cell gives the mean estimate of M(t) at time t=1, where $M(t)=\int_0^t \bar{F}(s)ds$. The bottom part of each cell gives the standard deviation of the estimate. For service time distributions other than exponential, a parametric estimate based on an erroneous exponential model is also given. The estimate in the case of an exponential model is $[1-EXP\{-t/\mu\}]$ where μ is the mean service time. For each service time distribution, the standard deviation of the parametric estimate of M(t) is smaller than that of the nonparametric estimate of M(t). That is, the efficiency of the parametric estimation method is better than the efficiency of the nonparametric estimation method. However, the results of a parametric fit assuming an erroneous exponential model show the worst performance. The true value of M(t) is not included within plus or minus three standard deviations of the erroneous estimate M(t). In the table, the nonparametric estimation methods seem to perform well in all cases with the cost of an inflation of variance. Hence the nonparametric estimation method is to be preferred when the service distribution is unknown.

To illustrate the efficiency of the nonparametric estimation methods, we simulated two possible ways to construct the approximate confidence interval for $M_N(t)$ for the bootstrap and the jackknife methods. Those ways are presented in chapter IV. For the jackknife estimation method, one procedure was to construct the confidence interval with the regular degrees of freedom, n-1, and the other used the reduced degrees of freedom, which is the number of different pseudo values. For the bootstrap estimation method, one way used the percentile method by the Monte Carlo process, and the other used the bias-corrected percentile method; there were 500 bootstrap replications. Nominal 68%, 80%, and 90% confidence intervals were constructed for each replication using each method. It was noted whether the confidence interval formed by a given method covered the true value M(t). The entire process was independently replicated with R=300 times. From these R replications we computed, for each method, the proportion $\hat{p}$ of the R confidence intervals which contained M(t), as well as the average length of the confidence intervals. If a method was performing adequately, $\hat{p}$ should be near $1-\alpha$, and a small mean length is desirable.

Tables IV to VII show the simulation results of several confidence interval procedures for four types of service time distribution; the exponential, the mixed exponential, the gamma, and the lognormal. The arrival process is Poisson with known arrival rate $\lambda=1$.

In order to compare the performance of these procedures to the normal confidence interval procedure, simulations were conducted, and nominal 68%, 80%, and 90% confidence limits were constructed for time t=1 for each replication. The normal confidence interval procedure is based on the order statistics of the service times. By the central limit theorem, the distribution of $M_N(t)$ is asymptotically normal

TABLE IV

COVERAGE AND LENGTH OF 100(1-α)% C. I
FOR EXPONENTIAL WITH μ=2, λ =1 AT T=1

		68 %		80 %		90 %	
		Length (s.d)	C. R.	Length (s.d)	C. R.	Length (s.d)	C. R.
Normal C. I. Procedure		0.0888 (0.0089)	19.33 69.33 11.33	0.1147 (0.0101)	12.00 83.67 4.33	0.1477 (0.0141)	4.67 92.00 3.30
Jack-knife	Reduced d. f	0.0911 (0.0089)	17.67 71.33 11.00	0.1187 (0.0102)	11.00 85.33 3.67	0.1548 (0.0142)	3.33 94.00 2.67
	Regular d. f	0.0897 (0.0090)	18.38 70.67 11.00	0.1163 (0.0103)	11.33 84.33 4.33	0.1505 (0.0144)	4.00 93.00 3.00
Boot-strap	Percen-tile method	0.0884 (0.0098)	18.67 69.00 12.33	0.1132 (0.0112)	12.00 82.00 6.00	0.1462 (0.0154)	4.33 92.00 4.67
	Bias-correct method	0.0887 (0.0098)	16.67 71.00 12.33	0.1137 (0.0112)	10.67 83.00 6.33	0.1467 (0.0154)	2.67 92.67 4.67

distributed as the number of data points n $\to\infty$. Thus, the 100(1-α)% normal confidence interval is given by

$$M_N(t) \pm z_{1-\frac{\alpha}{2}}\sqrt{Var[M_N(t)]} \qquad (5.1)$$

where $z_{1-\frac{\alpha}{2}}$ is the upper $1-\frac{\alpha}{2}$ point of the standard normal distribution and $Var[M_N(t)]$ is given by equation A.14 in appendix A.

Each cell in the tables contain the average and standard deviation of confidence interval length; and the proportion of intervals that are too high, (e.g. M (t)<L), where L is the lower bound of interval; the proportion of intervals covering the true value M(t), p; the proportion of interval that are too low, (e.g. M(t)>U), where U is the upper bound of interval. Table IV is for the exponential service time case with with μ =2; Table V is for the mixed exponential service time case with μ_1=2, μ_2=0.75, and P_1 =0.2; Table VI

TABLE V

COVERAGE AND LENGTH OF 100(1-α)% C.I. FOR
MIXED EXPONENTIAL WITH $\mu_1=2$, $\mu_2=.75$, $p_1=.2$, $\lambda=1$ AT T=1

		68 %		80 %		90 %	
		Length (s.d)	C. R.	Length (s.d)	C. R.	Length (s.d)	C.R.
Normal C.I. Procedure		0.0914 (0.0084)	14.67 71.67 13.67	0.1169 (0.0102)	10.00 83.00 7.00	0.1509 (0.0143)	5.00 90.67 4.33
Jack-knife	Reduced d.f	0.0936 (0.0085)	14.00 73.00 13.00	0.1208 (0.0103)	9.67 83.33 7.00	0.1581 (0.0143)	4.33 91.67 4.00
	Regular d.f	0.0923 (0.0085)	14.67 72.00 13.33	0.1185 (0.0103)	10.00 83.00 7.00	0.1538 (0.0145)	4.33 91.33 4.33
Boot-strap	Percen-tile method	0.0911 (0.0094)	14.33 71.67 14.00	0.1156 (0.0112)	11.33 81.33 7.33	0.1490 (0.0153)	4.33 89.67 4.33
	Bias-correct method	0.0913 (0.0094)	14.00 71.00 15.00	0.1161 (0.0112)	9.00 83.00 8.00	0.1495 (0.0153)	4.00 89.67 6.33

is for the gamma service time case with β=1 and κ=2; and Table VII is for the lognormal service time case with ξ=0.193 and δ^2=1.

The overall examination of the tabulations of confidence limit coverage and also the average and standard deviation of confidence interval length suggest that the bootstrap procedure is slightly better than the jackknife procedure; however, the difference is negligible. The normal confidence interval is also about the same as the jackknife and bootstrap procedures indicating that a sample size of 50 is large enough for the central limit theorem approximation to be adequate. All procedures produce almost the same average length of confidence interval with a good coverage rate, which falls within $\pm 2\sqrt{\frac{\alpha(1-\alpha)}{300}}$ of $1-\alpha$. Although the method of the reduced degrees of freedom used in the jackknife and the bias-correct percentile method applied in the bootstrap improved the coverage rate, the variance was

TABLE VI

COVERAGE AND LENGTH OF 100(1-α)% C.I.
FOR GAMMA WITH β=1, K=2, λ=1 AT T=1

		68 %		80 %		90 %	
		Length (s.d)	C. R.	Length (s.d)	C. R.	Length (s.d)	C.R.
Normal C.I. Procedure		0.0595 (0.0100)	22.00 68.33 9.67	0.0774 (0.0120)	12.33 80.33 7.33	0.0989 (0.0174)	10.67 86.33 3.00
Jack-knife	Reduced d.f	0.0617 (0.0102)	21.00 70.00 9.00	0.0813 (0.0122)	11.33 82.67 6.00	0.1062 (0.0178)	8.33 89.00 2.67
Jack-knife	Regular d.f	0.0601 (0.0102)	21.67 69.33 9.00	0.0785 (0.0121)	12.33 80.67 7.00	0.1008 (0.0177)	10.33 86.67 4.00
Boot-strap	Percen-tile method	0.0589 (0.0091)	21.33 69.97 9.00	0.0766 (0.0122)	12.00 80.33 7.67	0.0981 (0.0179)	9.33 86.67 3.00
Boot-strap	Bias-correct method	0.0595 (0.0100)	19.33 69.33 11.33	0.0777 (0.0121)	10.67 81.00 8.33	0.0990 (0.0180)	8.67 87.00 4.33

inflated. Furthermore, the amount of improvement was small and not significant. Hence, the original procedures for constructing the confidence interval for the jackknife and bootstrap are preferred in this case. Note that the coverage rates are skewed left slightly but almost balanced. It is a reason that the normal confidence interval procedure performs well.

Results will now be reported for the simulation of the case in which the arrival rate of the Poisson process is also unknown and must be estimated from interarrival time data. More computations are required for this case; however, the procedure is same. Each replication of the simulation generated 50 independent service times and 50 independent exponential interarrival times having mean 1. Confidence intervals were computed using both separated and paired jackknife procedures and the percentile method for the bootstrap. The number of bootstrap replications was

TABLE VII

COVERAGE AND LENGTH OF $100(1-\alpha)\%$ C.I.
FOR LOGNORMAL WITH $\xi=.193$, $\sigma^2=1$, $\lambda=1$ AT T=1

		68 %		80 %		90 %	
		Length (s.d)	C. R.	Length (s.d)	C. R.	Length (s.d)	C. R.
Normal C.I. Procedure		0.0769 (0.0075)	16.67 71.00 12.33	0.1007 (0.0096)	10.00 78.00 12.00	0.1272 (0.0126)	6.67 89.33 4.00
Jack-knife	Reduced d.f	0.0787 (0.0076)	16.33 71.33 12.33	0.1208 (0.0097)	9.67 79.67 10.67	0.1330 (0.0127)	6.33 89.67 4.00
	Regular d.f	0.0763 (0.0076)	16.67 70.33 13.00	0.1021 (0.0097)	10.00 79.00 11.00	0.1297 (0.0128)	6.67 89.33 4.00
Boot-strap	Percen-tile method	0.0763 (0.0084)	16.67 69.33 14.00	0.0998 (0.0105)	10.33 77.00 12.67	0.1258 (0.0133)	7.33 88.67 5.00
	Bias-correct method	0.0768 (0.0084)	16.00 68.33 15.67	0.1004 (0.0106)	8.67 78.33 13.00	0.1263 (0.0133)	6.33 88.67 5.00

1000. Nominal 68%, 80%, and 90% confidence limits were computed for each replication. The simulation was replicated 300 times.

Tables VIII to X report the results of the simulation. The quantities in the left part of each cell are the average and standard deviation (within parenthesis) of coverage interval length. The right part of each cell contains three quantities; the top value is the proportion of intervals that are too high; the center value is the proportion of intervals that cover the true value, $\hat{p}$; and the bottom part is the proportion of intervals that are too low.

In Table VIII (the case of 68% C.I.), the average length from the bootstrap shows outstanding performance with a small value of standard deviation. The paired jackknife procedure performs as well as the bootstrap procedure. This procedure reduced the standard deviation by more than half of that in the separated jackknife procedure, and also

improved the coverage rate. From the results of coverage rate in the table, it can be recognized immediately that the jackknife estimate, regardless of the application method, is often too low, while the bootstrap estimate tends may a little too high but is almost balanced in the number of confidence intervals that are too high or too low. It is the reason that the bias-corrected percentile method was not required in this case.

In this simulation, all the coverage rates fall within $\pm 2\sqrt{\frac{\alpha(1-\alpha)}{300}}$ of $1-\alpha$, (62.61, 73.38). Note that the average length of the confidence interval in the gamma service time case is the highest. When the arrival rate was known, the gamma service time case had the smallest average length. This indicates that the variability of the estimated arrival rate may be the dominate effect in the width of the confidence interval.

The results of Tables IX and X support the facts of discussion about Table VIII, This is the reasonable since the same random numbers were used to compute these confidence interval. The presentation of Table IX is exactly the same as the case of table VIII. All the coverage rate are fall within (75.38, 84.61), though the value of coverage rates fluctuate over the service time distribution cases. Obviously the paired jackknife procedure performs very well. The bootstrap procedure still has the best performance; however the value of coverage rate fluctuates greatly for the different service time distributions. For the 90% confidence interval case reported in Table X, some coverage rate fall outside the range (86.53, 93.46). The separated jackknife procedure produces low coverage rates outside of (86.53, 93.46), except for the exponential service time distribution. The paired jackknife procedure improved the coverage rate tremendously. Although the average lengths in the paired

jackknife procedure are slightly bigger than those in the bootstrap procedure, the overall performance is better than the bootstrap. Furthermore, the procedure in the case of gamma service time distribution, the bootstrap produced one coverage rate outside of (86.53, 93.46). However, this could be due to sampling fluctuation.

In general, all the confidence interval procedures performed very well for the exponential service time case, regardless of the level of the confidence interval. The procedures also worked well in general to produce 68% and 80% confidence interval. However, performance was more variable in the 90% confidence interval case. In most cases, the average length produced by the bootstrap procedure is the smallest, but the value of the coverage rate fluctuates for different service time distribution. The overall examination of the tabulations suggests that the paired jackknife procedure performs very well compared to the separated jackknife procedure and in some cases shows better performance than the bootstrap procedure.

TABLE VIII

COVERAGE AND LENGTH OF 68% C.I.
WITH UNKNOWN ARRIVAL RATE (N=50, R=300, B=1000)

	Jackknife				Bootstrap	
	Separated		Paired		Percentile	
	Length (s.d)	C. R.	Length (s.d)	C. R.	Length (s.d)	C.R.
Exponential	0.2674	7.00	0.2525	9.67	0.2436	18.33
		70.00		70.33		66.67
(μ = 2)	(0.1282)	23.00	(0.0570)	20.00	(0.0510)	15.00
Mixed exponen	0.1999	12.00	0.2077	15.00	0.2009	21.67
(μ_1 =2, μ_2=.75		65.67		68.00		64.00
P_1 =.2)	(0.0832)	22.33	(0.0429)	17.00	(0.0398)	14.33
Gamma	0.2917	10.00	0.2746	13.33	0.2632	21.67
		68.00		67.67		66.00
(β=1, k=2)	(0.1376)	22.00	(0.0599)	19.00	(0.0557)	12.33
Lognormal	0.2533	7.67	0.2513	10.00	0.2415	19.33
		72.00		72.33		69.67
(ξ=.193, σ^2=1)	(0.0998)	20.33	(0.0486)	17.67	(0.0461)	11.00

TABLE IX

COVERAGE AND LENGTH OF 80% C.I.
WITH UNKNOWN ARRIVAL RATE (N=50, R=300, B=1000)

	Jackknife				Bootstrap	
	Separated		Paired		Percentile	
	Length (s.d)	C. R.	Length (s.d)	C. R.	Length (s.d)	C.R.
Exponential	0.3447	5.33	0.3282	8.00	0.3162	12.67
		79.67		82.00		83.00
(μ = 2)	(0.1329)	15.00	(0.0641)	10.00	(0.0585)	4.33
Mixed exponen	0.2558		0.2688	5.67	0.2553	12.67
(μ_1 =2, μ_2=.75		75.67		82.33		83.00
P_1 =.2)	(0.1088)		(0.0569)	12.00	(0.0490)	4.33
Gamma	0.3646	5.00	0.3504	6.67	0.3375	16.33
		78.67		81.67		75.67
(β =1, k=2)	(0.1509)	19.33	(0.0703)	11.67	(0.0657)	8.00
Lognormal	0.3259	4.67	0.3231	6.33	0.3115	14.67
		77.67		79.33		75.67
(ξ=.193, σ^2=1)	(0.1279)	16.67	(0.0624)	14.33	(0.0585)	10.00

TABLE X

COVERAGE AND LENGTH OF 90% C. I.
WITH UNKNOWN ARRIVAL RATE (N=50, R=300, B=1000)

	Jackknife				Bootstrap	
	Separated		Paired		Percentile	
	Length (s.d)	C. R.	Length (s.d)	C. R.	Length (s.d)	C. R.
Exponential ($\mu = 2$)	0.4464 (0.1680)	1.33 90.00 8.67	0.4231 (0.0844)	2.67 92.33 5.00	0.4091 (0.0761)	8.67 89.00 2.33
Mixed exponen ($\mu_1 =2$, $\mu_2=.75$ $P_1 =.2$)	0.3199 (0.1213)	0.67 83.33 16.00	0.3387 (0.0679)	2.33 88.00 9.67	0.3301 (0.0609)	8.00 86.67 5.33
Gamma ($\beta=1$, k=2)	0.4891 (0.2321)	2.00 85.33 12.67	0.4586 (0.1053)	3.33 88.67 8.00	0.4416 (0.0959)	9.00 85.67 5.33
Lognormal ($\xi=.193$, $\sigma^2=1$)	0.4371 (0.1974)	1.67 85.33 13.33	0.4228 (0.0920)	3.00 90.00 7.00	0.4074 (0.0847)	7.00 88.67 4.33

VI. SUMMARY AND CONCLUSIONS

This thesis consider the problem of estimating M(t), the mean number of customers being served at time t for an M/G/∞ queue, using service time and interarrival time data. It is assumed that there are no customers being served at time 0. Two cases are considered. In one the parametric form of the service time distribution is assumed known. In this case M(t) is a function of the estimated parameters. In the situation in which the arrival rate of the Poisson process is also assumed known and the parametric form of the service time distribution is exponential, approximation to the bias and variance of the estimate are derived. Further, simulation is used to study a normal confidence interval procedure.

For the other case the parametric form of the service time distribution is unknown. The empirical distribution of the service time distribution is used in the estimate of M(t). In the situation in which the arrival rate λ is assumed known, the distribution of the estimate is derived in Appendix A. The bootstrap and jackknife estimates with known are studied in Appendix B and C. Simulation was used to assess the performance of confidence interval procedures using a normal approximation the jackknife and the bootstrap. The simulation results for the case in which the arrival rate λ is known indicate that:

(1) The parametric estimation method appears the most powful method when the parametric assumption is correct, but the performance is seriously degraded if the assumption is not appropriate.

(2) When an erroneous parametric (exponential) model is assumed, the initial estimates of mean number in service are poor. However, as $t \to \infty$, the erroneous parametric estimate approaches the same value as the other estimates. This is because as $t \to \infty$ all the estimates approach the sample mean of the service time data.

(3) The estimate obtained by using the empirical distribution is unbiased with a larger variance than a parametric estimate based on a correct model.

(4) Simulation results indicate there is not much difference between jackknife and bootstrap confidence interval procedures.

(5) The nonparametric normal confidence interval procedure performs as well as the procedure in (4) since the distribution of the estimate is almost symmetric. The improvement by the use of adjusted degrees of freedom in the jackknife and the bias-corrected percentile in the bootstrap is small.

We now discuss the simulation results for the case in which the arrival rate for the Poisson process is also unknown and is estimated using interarrival time data. The service times are generated from four types of distribution. The percentile method for the bootstrap and paired and separated techniques for the jackknife were used to construct the confidence intervals. Tables I and II, which are the results of a parametric confidence interval procedure in chapter III, are compared with the results of the nonparametric confidence interval procedures. The simulation results indicate that:

(1) The nonparametric confidence interval procedure works as well as the parametric case, even though the length of the confidence interval is wider than the parametric one.

(2) In the overall examination, the percentile method of the bootstrap shows the best performance. The paired jackknife procedure also has similiar results to the bootstrap approach. The results of these two nonparametric procedures show the almost same level of performance with the parametric one. However, the separated jackknife procedure produces poor results.

(3) The results of the jackknife procedures produce intervals that are always biased upward. Efron (1981) reported similiar results. [Ref. 13]

(4) Since the bootstrap procedures require a large amount of computation, the jackknife is the method of choice if one does not want to do the bootstrap computations.

In general, the nonparametric methods of the bootstrap and the jackknife performs very well, regardless the complexing of the estimation problem. Of cource, if the parametric estimation method can be applied, the results are clearly superior. However, the application of the

parametric estimation is a highly limited because the parametric assumption is often difficult to verify. When the estimate is simple enough, which is the nonparametric estimate when the arrival rate is known, and the asymptotic distribution of estimate can be obtained, the nonparametric normal confidence interval procedure performs well, and more complicate computations such as the jackknife and the bootstrap method are not required. However, the jackknife and the bootstrap method have a good performance for the more complicated problem in which the arrival rate is unknown. The bootstrap confidence intervals show the best performance but the paired jackknife procedure achieve the same level of performance with less computation than the bootstrap in this problem.

APPENDIX A

CALCULATING THE BIAS AND THE VARIANCE OF $M_N(T)$ WITH KNOWN ARRIVAL RATE

It will be assume that the rate of Poisson process λ is known and is equal to 1. The nonparametric estimate $M_N(t)$ is given by

$$M_N(t) = \int_0^t [1-F(s)]ds \qquad (A.1)$$

Using the empirical cumulative density function F_n, the nonparametric estimate is

$$M_N(t) = \frac{1}{n}\sum_{i=1}^{\hat{K}} s_{(i)} + \frac{n-\hat{K}}{n}\, t \qquad (A.2)$$

where the observation $S_{(i)}$'s are the order statistics of the indeprndent and identically distributed service times with unknown distribution F. To find the distribution of $M_N(t)$, we will study the distribution of $\{S_{(i)}\}$.

Let $X_1, X_2, \ldots, X_n$ be independent, identically distributed random variable with distribution function F. Let N be the number of X_i's which are less than t. Let $X_{(i)}$ denote the i^{th} smallest X_i. By the definition of conditional probability,

$$P\{X_{(1)} \leq x | N_t=1\} = \frac{P\{X_{(1)} \leq x,\ N_t=1\}}{P\{N_t=1\}} \qquad (A.3)$$

for x<t. Since the random variable N_t has a binomial distribution with a parameter F(t), we can rewrite the equation A.3 to obtain

$$P\{X_{(1)} \leq x | N_t=1\} = \frac{\binom{n}{1} F(x) [\bar{F}(t)]^{n-1}}{\binom{n}{1} F(t) [\bar{F}(t)]^{n-1}}$$

$$= \frac{F(x)}{F(t)} \qquad (A.4)$$

Let $f(x)$ be the density of the distribution function F. The conditional probability given $N_t=2$,

$$P\{x_1 \leq X_{(1)} \leq x_1 + dx_1 ,\ x_2 \leq X_{(2)} \leq x_2 + dx_2 \mid N_t = 2\}$$

$$= 2 \frac{f(x_1)dx_1}{F(t)} \frac{f(x_2)dx_2}{F(t)} \qquad (A.5)$$

for $x_1 < x_2$.

Given $N_t=\hat{K}$, the conditional distribution of the values of the unordered X_i that lie in $(0,t]$ is that of independent random variables with distribution function $\frac{F(x)}{F(t)}$, for $0 \leq x \leq t$. Thus, given $N_t=\hat{K}$, $M_N(t)$ has the same distribution as a constant plus the sum of $\hat{K}$ independent identically distributed random variables. Thus the expectation of $M_N(T)$ can computed by the property of conditional expectation,

$$E[M_N(t)] = E[E[M_N(t)|N_t]] \qquad (A.6)$$

Given $N_t=\hat{K}$,

$$E[M_N(t)|N_t=\hat{K}] = \frac{\hat{K}}{n} \int_0^t x \frac{F(dx)}{F(t)} + \frac{n-\hat{K}}{n} t \qquad (A.7)$$

Since the random variable N_t has a binomial distribution with the parameter $F(t)$,

$$E[M_N(t)] = \frac{nF(t)}{n} \int_0^t \frac{F(dx)}{F(t)} + \frac{n-nF(t)}{n} t$$

$$= \int_0^t xF(dx) + [1-F(t)]t \qquad (A.8)$$

To check the bias of the estimate $M_N(t)$, using the integration by part of equation A.1, the true estimate is given by

$$M_N(t) = t[1-F(t)] + \int_0^t tF(dx) \qquad (A.9)$$

Thus the estimate $M_N(t)$ is unbiased. Using conditional expectations, the variance is computed by

$$Var[M_N(t)] = E[(M_N(t) - E[M_N(t)])^2]$$

$$= E[(M_N(t) - E[M_N(t)|N_t])^2]$$

$$+ 2E[(M_N(t) - E[M_N(t)|N_t])(E[M_N(t)|N_t] - E[M_N(t)])]$$

$$+ E[(E[M_N(t)|N_t] - E[M_N(t)])^2] \qquad (A.10)$$

Computing each of the individual terms, we obtain

$$E[(M_N(t) - E[M_N(t)|N_t])^2]$$

$$= \frac{F(t)}{\eta} \{ \int_0^t x^2 \frac{F(dx)}{F(t)} - [\int_0^t x \frac{F(dx)}{F(t)}]^2 \} \qquad (A.11)$$

and

$$E[(E[M_N(t)|N_t] - E[M_N(t)])^2]$$

$$= \frac{1}{\eta} F(t)\bar{F}(t)[\int_0^t x \frac{F(dx)}{F(t)} - t]^2 \qquad (A.12)$$

where $\bar{F}$ is the survival distribution function of F. The second term of right term of equation A.10 turn out to be zero. Thus the variance of $M_N(t)$ is obtained by sum of two equations. The resulting variance is

$$Var[M_N(t)] = \frac{1}{\eta} \{ \int_0^t x^2 F(dx) - [\int_0^t xF(dx)]^2 + t^2 F(t)\bar{F}(t)$$

$$- 2t\bar{F}(t)[\int_0^t xF(dx)] \} \qquad (A.13)$$

Notice that the variance estimate of $M_N(t)$ is equal to $\frac{1}{\eta}Var(X)$ as $t \to \infty$. A nonparametric estimate of $Var[M_N(t)]$ is

$$\widehat{Var}[M_N(t)] = \frac{1}{\eta}\{\frac{1}{\eta}\sum_{i=1}^{\hat{K}} s_{(i)}^2 - [\frac{1}{\eta}\sum_{i=1}^{\hat{K}} s_{(i)}]^2 + t^2\frac{\hat{K}}{\eta}[1-\frac{\hat{K}}{\eta}]$$

$$- 2t \frac{\hat{K}}{n} \left[\frac{1}{n} \sum_{i=1}^{\hat{K}} s_{\hat{i}} \right] \} \qquad \text{(A.14)}$$

where $\hat{K}$ is the number of service times that are less than t.

APPENDIX B

JACKKNIFE ESTIMATE OF NONPARAMETRIC ESTIMATE

It will be assumed that the arrival rate λ is known and equal to 1. The nonparametric estimate is given by

$$\hat{M}_N(t) = \frac{1}{n}\sum_{i=1}^{\hat{K}} s_{(i)} + \frac{n-\hat{K}}{n}\, t \tag{B.1}$$

where $S_{(i)}$'s are the order statistics of the service times and assumed that the random variable $\hat{K}$ exists such that $S_{(1)} \leq S_{(2)} \ldots \leq S_{(K)} \leq t \leq S_{(K+1)} \leq \ldots \leq S_{(n)}$. The estimate $\hat{M}^{\,i}_{n-1}(t)$ for the data set obtained by deleting the i^{th} point from the sample is

$$\hat{M}^{\,i}_{n-1}(t) = \begin{cases} \dfrac{1}{n-1}\sum_{j=1}^{\hat{K}} s_{(j)} + \dfrac{(n-1)-\hat{K}}{n-1}\, t & \text{if } i > K \\[2ex] \dfrac{1}{n-1}\sum_{\substack{j=1 \\ j\neq i}}^{\hat{K}} s_{(j)} + \dfrac{n-\hat{K}}{n-1}\, t & \text{if } i \leq K \end{cases} \tag{B.2}$$

The pseudo-value $\hat{M}_i(t)$ is computed by

$$\hat{M}_i(t) = n\hat{M}_{all}(t) - (n-1)\hat{M}^{\,i}_{n-1}(t) \tag{B.3}$$

where $\hat{M}_{all}(t)$ is the estimate of $M_N(t)$ based on all the data. substituting the estimate $\hat{M}^{\,i}_{n-1}(t)$ in equation B.3, we get

$$\hat{M}_i(t) = \begin{cases} S_{(i)} & \text{if } i \leq K \\[2ex] t & \text{if } i > K \end{cases} \tag{B.4}$$

Since the jackknife estimate is the average of the pseudo-values, the estimate is given by

$$\hat{M}_J(t) = \frac{1}{n}\sum_{i=1}^{\hat{K}} s_{(i)} + \frac{n-\hat{K}}{n}\, t \tag{B.5}$$

Thus, the jackknife estimate is the same as the original nonparametric estimate of $M_N(t)$. The jackknife estimate of variance is

$$\hat{Var}[M_J(t)] = \frac{1}{n-1}[\frac{1}{n}\sum_{i=1}^{n}\{\hat{M}_i(t)\}^2 - \{\frac{1}{n}\sum_{i=1}^{n}\hat{M}_i(t)\}^2\} \qquad (B.6)$$

where

$$\sum_{i=1}^{n}\{\hat{M}_i(t)\}^2 = \sum_{i=1}^{\hat{K}}s_{(i)}^2 - (n-\hat{K})t^2$$

and

$$\{\sum_{i=1}^{\hat{K}}\hat{M}_i(t)\}^2 = [\sum_{i=1}^{\hat{K}}s_{(i)} + (n-\hat{K})t]^2$$

Thus the variance can be rewritten as

$$\hat{Var}[M_J(t)] = \frac{1}{n-1}\{\frac{1}{n}\sum_{i=1}^{\hat{K}}s_{(i)}^2 - [\frac{1}{n}\sum_{i=1}^{\hat{K}}s_{(i)}]^2 + t^2\frac{n-\hat{K}}{n}\ \frac{\hat{K}}{n}]$$

$$- 2t\ \frac{n-\hat{K}}{n}[\frac{1}{n}\sum_{i=1}^{\hat{K}}s_{(i)}]\} \qquad (B.7)$$

Comparing equation B.7 with equation A.14, it is seen that the jackknife estimate of variance is greater than the estimate of equation A.14 by a multiplicative constant $\frac{n}{n-1}$.

APPENDIX C

CONDITIONAL DISTRIBUTION OF BOOTSTRAP ESTIMATE

It will be assumed that the arrival rate λ is known. and is equal to 1. Let $S_1, S_2, .., S_n$ denote random service times from a CDF F, and let $S_{(1)} \leq S_{(2)} \leq .. \leq S_{(K)} \leq t \leq S_{(K+1)} \leq .. \leq S_{(n)}$ denote the corresponding order statistics. The nonparametric estimate of $\int_0^t \bar{F}(s)ds$ is

$$\hat{M}_N(t) = \frac{1}{n} \sum_{i=1}^{\hat{K}} s_{(i)} + \frac{n-\hat{K}}{n} t \qquad (C.1)$$

Let B_i, i=1 to n, be independent random variables having the same distribution as draws with replacement from ($s_1, s_2, .., s_n$), and let $b_{(i)}$, i=1 to n, be the corresponding order statistics. A bootstrap realization of the nonparametric estimate is

$$\hat{M}_B(t) = \frac{1}{n} \sum_{b_{(i)} \leq t} b_{(i)} + \frac{1}{n}[n - \sum_{i=1}^{n} I(b_{(i)} \leq t)]t \qquad (C.2)$$

where

$$I(x \leq t) = \begin{cases} 1 & \text{if } x \leq t \\ 0 & \text{otherwise} \end{cases}$$

To compute the distribution of $\hat{M}_B(t)$, the Laplace transform is used [Ref. 12]. The Laplace transform of $\hat{M}_B(t)$ is

$$E[\exp(-\S \hat{M}_B(T))] = E[E[\exp(-\S M_B(t))|N_t]] \qquad (C.3)$$

by the property of conditional expectations, where N_t is the number of bootstrap samples which are less than or equal to t. We compute the right hand side of equation C.3 separately. First

$E[\exp(-\S \hat{M}_B(t))|N_t = 1]$

$$= \exp(-\xi t)[\exp(-\xi \frac{t}{n}) \sum_{i=1}^{\hat{K}} \frac{1}{\hat{K}} \exp(-\xi \frac{S_{(i)}}{n})]^{\ell} \quad (C.4)$$

is computed, where the random variable N_t has a binomial distribution with the parameter $F(t) = \frac{\hat{K}}{n}$. Thus, from equation C.3 the Laplace transform of $\hat{M}_B(t)$ is

$$E[\exp(- M(t)] = \exp(-\xi t) \sum_{\ell=1}^{\hat{K}} [\exp(-\xi \frac{t}{n}) \sum_{i=1}^{\hat{K}} \frac{1}{\hat{K}} \exp(-\xi \frac{S_{(i)}}{n})]^{\ell} \binom{n}{\ell} (\frac{\hat{K}}{n})^{\ell} (\frac{n-\hat{K}}{n})^{n-\ell}$$

$$= \exp(-\xi t)[\frac{\hat{K}}{n} \exp(\xi \frac{t}{n}) \sum_{i=1}^{\hat{K}} \frac{1}{\hat{K}} \exp(-\xi \frac{S_{(i)}}{n}) + \frac{n-\hat{K}}{n}]^{n} \quad (C.5)$$

Let us define a random variable Y having the following distribution

$$Y = \begin{cases} \frac{S_{(i)}}{n} & \text{w.p } \frac{1}{n} \quad \text{if } i \leq \hat{K} \\ \frac{t}{n} & \text{w.p } \frac{n-\hat{K}}{n} \quad \text{if } i > \hat{K} \end{cases} \quad (C.6)$$

Then the Laplace transform of Y is

$$E[\exp(-\xi Y)] = \frac{1}{n} \sum_{i=1}^{\hat{K}} \exp(-\xi \frac{S_{(i)}}{n}) + \frac{n-\hat{K}}{n} \exp(-\xi \frac{t}{n}) \quad (C.7)$$

Thus, $\hat{M}_B(t)$ has the same distribution as the sum of n independent random variables having the same distribution as Y. For the fixed time t, given the order statistics, $S_{(1)} < S_{(2)} < .. < S_{(K)} < t < S_{(K+1)} < .. < S_{(n)}$, the expectation of $\hat{M}_B(t)$ is written by

$$E[\hat{M}_B(t) | \text{data}] = nE[Y | \text{data}]$$

$$= \frac{1}{n} \sum_{i=1}^{\hat{K}} s_{(i)} + \frac{n-\hat{K}}{n} t \quad (C.8)$$

Thus, the bootstrap estimate is asymptotically an unbiased estimate of $M_N(t)$. The variance is

$$\text{Var}[\hat{M}_B(t) | \text{data}] = n\text{Var}[Y] \quad (C.9)$$

The variance of Y can be derived using the equation C.6 Since

$$E[Y^2] = \frac{1}{n}\sum_{i=1}^{\hat{K}}\left[\frac{S_{(i)}}{n}\right]^2 + \frac{n-\hat{K}}{n}\left(\frac{t}{n}\right)^2 \qquad \text{(C.10)}$$

Hence the asymptotic bootstrap variance estimate of $M_N(t)$ is given by

$$\hat{Var}[M_B(t)|\text{data}] = \frac{1}{n}\left\{\frac{1}{n}\sum_{i=1}^{\hat{K}} s_{(i)}^2 - \left[\frac{1}{n}\sum_{i=1}^{\hat{K}} s_{(i)}\right]^2 + t^2\frac{\hat{K}}{n}\cdot\frac{n-\hat{K}}{n}\right.$$

$$\left. - 2t\,\frac{n-\hat{K}}{n}\left[\frac{1}{n}\sum_{i=1}^{\hat{K}} s_{(i)}\right]^2\right\} \qquad \text{(C.11)}$$

That is, the asymptotic bootstrap estimate of variance is the same as the nonparametric estimate (equation A.14).

LIST OF REFERENCES

1. Gross, D. and Harris, C.M., Fundamentals of Queueing Theory, John Wiley and Sons, 1974.

2. Brown, M. and Ross, S.M., "Some Results for Infinite Server Queues", J. Appl. Prob., 6, pp. 604-611, 1969.

3. Takacs, L., Introduction to the Theory of Queues, Oxford University Press, 1962.

4. Cox, D. R., "Some Problems of Statistical Analysis Connected with Congestion", Proceedings of the Symposium on Congestion Theory, pp. 289-316, 1964.

5. Mirasol, N. M., "The Output of an M/G/ Queueing System is Poisson", Opns. Res., 11, pp. 282-284, 1963.

6. Lewis, P. A. W. and Uribe, L., "The New Naval Postgraduate School Random Number Package LLRANDOMII", NPS 55-81-005, Monterey, California, 1981.

7. Quenouille, M. H., "Notes an Bias in Estimation", Biometrika, 4, pp. 353-360, 1956.

8. Tukey, J. W., "Bias and Confidence in Not Quite Large Samples", Ann. Math. Statist., 29, p. 614, 1958.

9. Miller, R. G., "The Jackknife - a Review", Biometrika, 61, pp. 1-15, 1974.

10. Efron, B., "Bootstrap Methods : Another Look at the Jackknife", Ann. Statist., 7, pp. 1-26, 1979.

11. Efron, B., The Jackknife, the Bootstrap, and other Resampling Plans", Technical Report No. 63, Dept. of Statistics, Stanford University, 1980.

12. Feller, W. An Introduction to Probability Theory and its Applications, Vol. 2, New York, Wiley, 1966.

13. Efron, B.,"Nonparametric Standard Errors and Confidence Intervals", Canadian Journal of Statistics, 9, pp. 139-172, 1981.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22304-6145	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93943-5002	2
3.	Professor Patricia A. Jacobs, Code 55Jc Department of Operations Research Naval Postgraduate School Monterey, California 93943-5000	1
4.	Professor Donald A. Gaver, Code 55gv Department of Operations Research Naval Postgraduate School Monterey, California 93943-5000	2
5.	Professor Alan W. Washburn, Code 55 Department of Operations Research Naval Postgraduate School Monterey, California 93943-5000	1
6.	CPT. Choi, Eui Joung Devision of System Analysis P.O. Box 201-17 Youngdeongpo Gu Singil 7 Dong Seoul Korea	1
7.	LT. Park, Dong Keun Incheon si, Booku Sangok Dong 307 Bupyeong Hyeun Dai Apt 113-103 Seoul Korea	10
8.	LCDR. Kim, Won Bong 458 Ramona Ave. #1 Monterey, California 93940	1

1[illegible]7-197

217536

Thesis
P157404 Park
c.1

Parametric and non-parametric estimation of the mean number of customers in service for AN M/G/00 queue.

217536

Thesis
P157404 Park
c.1

Parametric and non-parametric estimation of the mean number of customers in service for AN M/G/00 queue.

thesP157404
Parametric and nonparametric estimation
3 2768 000 66022 9
DUDLEY KNOX LIBRARY